

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:24:22 ; Search time 24 Seconds
(without alignments)
809.130 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHHMESGESVSSNQS.....PAFIQVVLNIALSFGYLS 660

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1690	52.7	361	4	US-09-556-877-299 Sequence 299, App
2	1690	52.7	361	4	US-09-620-412C-299 Sequence 299, App
3	248.5	7.8	10182	4	US-09-134-001C-3159 Sequence 3159, Ap
4	237	7.4	2137	4	US-09-134-001C-4463 Sequence 4463, Ap
5	235	7.3	2285	4	US-09-308-375-2 Sequence 2, Appli
6	219	6.8	3696	4	US-09-134-001C-5080 Sequence 5080, Ap
7	194.5	6.1	1566	2	US-08-687-956A-23 Sequence 23, Appli
8	194	6.1	1093	5	PCT-US93-03077-1 Sequence 1, Appli
9	189.5	5.9	3248	1	US-08-353-700-1 Sequence 1, Appli
10	189.5	5.9	3248	5	PCT-US95-16216-1 Sequence 1, Appli
11	186	5.8	3788	4	US-09-336-447A-76 Sequence 76, Appli
12	183.5	5.7	955	2	US-08-428-414A-3 Sequence 3, Appli
13	183.5	5.7	1561	3	US-08-894-017-23 Sequence 23, Appli
14	183.5	5.7	2101	1	US-08-466-390-4 Sequence 4, Appli
15	183.5	5.7	2101	1	US-08-470-950-4 Sequence 4, Appli
16	183.5	5.7	2101	1	US-08-467-781-4 Sequence 4, Appli
17	183.5	5.7	2101	1	US-08-195-487-4 Sequence 4, Appli
18	183.5	5.7	2101	2	US-08-483-924-4 Sequence 4, Appli
19	183.5	5.7	2101	4	US-08-483-924-4 Sequence 4, Appli
20	183.5	5.7	2101	4	US-09-452-294-1 Sequence 1, Appli
21	182.5	5.7	955	1	PCT-US93-06160-4 Sequence 1, Appli
22	182.5	5.7	955	1	US-08-006-676B-1 Sequence 1, Appli
23	182.5	5.7	955	1	US-08-283-845-2 Sequence 1, Appli
24	182.5	5.7	955	5	PCT-US94-00324-1 Sequence 2, Appli
25	180	5.6	2482	4	US-09-336-447A-5 Sequence 1, Appli
26	178.5	5.6	1354	3	US-08-328-254-6 Sequence 5, Appli
27	175	5.5	1939	4	US-08-685-871-2 Sequence 6, Appli
					US-09-310-187A-1 Sequence 1, Appli

28	174	5.4	2411	4	US-09-268-347-36 Sequence 36, Appli
29	173.5	5.4	1886	4	US-08-938-105-3 Sequence 3, Appli
30	172.5	5.4	873	4	US-09-336-447A-13 Sequence 13, Appli
31	172	5.4	534	4	US-09-103-664A-2 Sequence 2, Appli
32	171	5.3	676	4	US-09-336-447A-15 Sequence 15, Appli
33	170.5	5.3	889	4	US-09-134-001C-4318 Sequence 4, Appli
34	170.5	5.3	3111	2	US-08-460-309-4 Sequence 4, Appli
35	170.5	5.3	3111	2	US-08-125-077-4 Sequence 4, Appli
36	169.5	5.3	1565	6	5352450-2 Patent No. 5352450
37	167	5.2	1073	4	US-09-541-782-6 Sequence 6, Appli
38	167	5.2	1073	4	US-09-723-820-6 Sequence 6, Appli
39	166.5	5.2	1786	4	US-08-973-462-8 Sequence 8, Appli
40	166.5	5.2	1912	1	US-08-409-995-4 Sequence 4, Appli
41	166.5	5.2	1912	3	US-08-685-467-4 Sequence 4, Appli
42	166.5	5.2	2353	4	US-09-377-155-33 Sequence 33, Appli
43	166.5	5.2	2353	4	US-08-913-942-4 Sequence 4, Appli
44	166.5	5.2	2353	4	US-09-669-974-33 Sequence 33, Appli
45	166.5	5.2	2354	4	US-09-268-347-47 Sequence 47, Appli

ALIGNMENTS

RESULT 1									
US-09-556-877-299									
Sequence 299, Application US/09556877									
Patent No. 6432916									
GENERAL INFORMATION:									
APPLICANT: Probst, Peter									
APPLICANT: Bhatia, Ajay									
APPLICANT: Skeiky, Yasir									
APPLICANT: Fling, Steve									
APPLICANT: Maisonneuve, Jeff									
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND									
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION									
FILE REFERENCE: 210121.469C5									
CURRENT APPLICATION NUMBER: US/09/556.877									
NUMBER OF SEQ ID NOS: 305									
SOFTWARE: FASTSEQ for Windows Version 3.0/4.0									
SEQ ID NO 299									
LENGTH: 361									
TYPE: PRT									
ORGANISM: Chlamydia									
US-09-556-877-299									
Query Match									
Best Local Similarity 97.0% Pred. No. 9e-109;									
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;									
Score 1690; DB 4; Length 361;									
QY	300	HTGLTDSPLVKKAEQISOAKDIOEIRPGSDPIYVPGSGASAGSAGALKSSNNG	359						
DB	1	HOEIDSPLVKKAEQIQNOADODIOTIPPGSDIPYVPGSGASAGSAGALKSSNNG	60						
QY	360	RSLILDDVDNEMAIATGQFRSMTEOFNNVNNPATAKELQMEQULTMSDOLVGADEL	419						
DB	61	RSLILDDVDNEMAIATGQFRSMTEOFNNVNNPATAKELQMEQULTMSDOLVGADEL	120						
QY	420	PAETQAIKDALQALQKPSADGLATAMQGVAFAAKVGSGAGTAVQVMVKKQLYKTA	479						
DB	121	PAETQAIKDALQALQKPSADGLATAMQGVAFAAKVGSGAGTAVQVMVKKQLYKTA	180						
QY	480	STSSSSSTAALSDGYSAKTKTLNLSYSSRGVQSAISQOTANPALSRVSSSGIESQ	539						
DB	181	STSSSSSTAALSDGYSAKTKTLNLSYSSRGVQSAISQOTANPALSRVSSSGIESQ	240						
QY	540	ADASORAAETTYRDSOTGDDYVSRLOVLDLSMTSTVSNPOANOETIMOKLTA	599						
DB	241	ADASORAAETTYRDSOTGDDYVSRLOVLDLSMTSTVSNPOANOETIMOKLTA	300						
QY	600	GYPVQNSADSLOKFAQLEREFEVDGERSLAESQENAFKQPAFTQVVLNIALS	659						
DB	301	GYPVQNSADSLOKFAQLEREFEVDGERSLAESQENAFKQPAFTQVVLNIALS	360						

OY 660 S 660
Db 361 S 361

RESULT 2
US-09-620-412C-299

; Sequence 299, Application US/09620412C
; Patent No. 6448234

; GENERAL INFORMATION:

; APPLICANT: Steven P. Fling

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND

; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

; FILE REFERENCE: 210121.469C7

; CURRENT APPLICATION NUMBER: US/09/620,412C

; CURRENT FILING DATE: 2000-07-20

; NUMBER OF SEQ ID NOS: 363

; SOFTWARE: FastSeq for Windows Version 3.0/4.0

; SEQ ID NO: 299

; LENGTH: 361

; TYPE: PRF

; ORGANISM: Chlamydia

US-09-620-412C-299

Query Match

Best Local Similarity 52.7%; Score 1690; DB 4; Length 361;
Best Local Similarity 97.0%; Pred. No. 9e-109;

Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 300 HTGLTSPLYKKAEEQISQAOKIOETKPSGSDIPYVPGSGSAASAGALKSSNNG 359
Db 1 HOBIASPLVKKAEEQINQAODIQITPESGLDPIVPGSAAASAGALKSSNNG 60
OY 360 RISLLDDVDNEMAATALOGFRSMIOFNVPATAKELDAMEAOLTAMSDOLVGADGEL 419
Db 61 RISLLDDVDNEMAATALOGFRSMIOFNVPATAKELDAMEAOLTAMSDOLVGADGEL 120
OY 420 PAEIQAIKDALQAOLKOPSDGLATAMGOVAFAAKVGSGSAGTACTVQNNVKOLYKTAF 479
Db 121 PAEIQAIKDALQAOLKOPSDGLATAMGOVAFAAKVGSGSAGTACTVQNNVKOLYKTAF 180
OY 480 SSS 539
Db 181 SSS 240
OY 540 ADSORAETIVDSOTGLGVYSLQVLDLMSITVSNPOANOBEIKOTLTAISKRAPOF 599
Db 241 ADSORAETIVDSOTGLGVYSLQVLDLMSITVSNPOANOBEIKOTLTAISKRAPOF 300
OY 600 GYPVONSADSLQKFAAQLEREFVDEGERSIAESQENAFKOPAFIOQVLVNIASLFSGYL 659
Db 301 GYPVONSADSLQKFAAQLEREFVDEGERSIAESQENAFKOPAFIOQVLVNIASLFSGYL 360
OY 660 S 660
Db 361 S 361

RESULT 3

US-09-134-001C-3159

; Sequence 3159, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRF
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159

Query Match

Best Local Similarity 7.8%; Score 248.5; DB 4; Length 10182;
Best Local Similarity 19.7%; Pred. No. 1.6e-07;
Matches 141; Conservative 128; Mismatches 256; Indels 189; Gaps 31;

OY 18 NQSSAMPILINGQIASNSEKREKASEASPSASSSSVSSPSSAKNALISLDAIKN 77
Db 6003 SSPTINQIANVAAYQAKSNLHGDKLOHDKDSAKQTTAOLQNLNSAK--HHEDSLDNE 6059
OY 78 SSPTD---SLSOLEASTSTVTRVAKDYDEAKS-----FDTAKSLENA 121
Db 6060 STRTQVQHDLEQAOLDLGMLAKESIKDTYINVSNGNTINAEPSKKQAYDAVQAQNI 6119
OY 122 KTLAEYET--KMDLMAALODMERLANS-----DPSNNHTEEVNNIKALEAOKD-- 169
Db 6120 INGTNPPTINKGNVTTATQTVKNTKDALDQDHRLEFAKNNAQOTIRNLNNAQDAEK 6179
OY 170 -----TIDKLNK-LVTLQONKSLTEVLTCTDSADQIPA-----INSOLETKNSADQI 217
Db 6180 NLVNSASTLEQVQONQOTQOOLDNMGELQSIJAKKDQVKAOSKYLNEEDPOIKQNDQAV 6239
OY 218 IKDLERQNTSYEAVLNTNAGEVIRKASEAGIKLGOALQSIYDAGDQSOAAVLQAOQNSPD 277
Db 6240 -----QVETIINETQNPBELKAN-----IDQATQVQNA-----EQALHGAKEKLNQK 6283
OY 278 NIATK-----ELIDAETRVNE-----LKO--EHGTLTSPLYKKAEEQISQA 319
Db 6284 QTSSTLEGLTDLTDQAREKREIQNTSRRDQIKQIQAKALNDA--MKLKEQV--A 6339
OY 320 QKDIQETKPSGSDIPYVPGSGSAASAGALKSS-----NNGRISLLDDVDNEMA 375
Db 6340 QKQGVH--ANSQY---TNEQSAQDAVNNALKQAEIITNNSNPNLNAQDITNALNNT 6392
OY 376 A-----LOGFRSMIEQFNVPATAKELQAMEAQLTAMSDOLVGADGELPAEIQAIKDAL 430
Db 6393 KQADNLHGAKQKQKQKNTTQALIGNLNLNQKQKALQATNGAITS-----RDQV 6443
OY 431 AQALKOPSDGLATAMGOVAFAAKVGSGSAGTACTVQNNVKOLYKTAFSSSSSSSYAA 490
Db 6444 AEKLEK--AEALDEAM-----KOLEQVQVODDQISNSPF 6476
OY 491 LSDGYSAKTLNLSYESRSGVQSAISQANPALSRVSRSIGESGRSADASQRAETI 550
Db 6477 INEDSDKQKT---YNDKIQAAKEIITQNTSPTLXK-----QKTAIDL 6515
OY 551 ---VRD-----SOTLGIVYSLQVLDL-----MSTIVSNPOANO-----EI 585
Db 6516 QNIKDAVNNLHGDKLQAKSQDANNQNLNHDLTREOKNHFKRLINNAQTRDREYNNQLEI 6575
OY 586 MOKLTASISKAPQFGYPVONSADSLQKFAAQLEREFVDEGERSIAESQENAFK 639
Db 6576 AKQNGDMSTLHK-----VINDKQIQIOLHS-----NYINADNDKQNYDNAIKE 6619

RESULT 4

US-09-134-001C-4463

; Sequence 4463, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

;; PRIOR APPLICATION NUMBER: US 60/055,779
;; PRIOR FILING DATE: 1997-08-14
;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 4463
;; LENGTH: 2137
;; TYPE: PR1
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4463

Query Match 7.48; Score 237; DB 4; Length 2137;
Best Local Similarity 19.98; Pred. No. 1.1e-07;
Matches 136; Conservative 134; Mismatches 316; Indels 96; Gaps 23;

```
QY 9 ESSESVSSNOGSMPIINGOIASNSE-----TKESTKASEA---SPSASSVSSWFLS 60
Db 1079 ESDSTSTSLSESTSTSLSGSTASSTSDASTSTSESDSTSESTSLSESLSTSVSDSTAS 1138
QY 61 SAKNALISLRDAIILKNSSPTDSLQLEASTSTSTVTTRAADYDEAKSNFDTAKSGLEN 120
Db 1139 TSESASTSTSES--ESNASTSLSGSLSTSLSDSTSTSTSTSTSTSTSTSTSTSTSTSTST 1190
QY 121 AKTAEIETKADLMALQDMERLANSPSNNHTEEVNNIKALEAQKDTIDKLNKLVTL 180
Db 1191 TSTSLSESTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSA 1245
QY 181 QNOKSLTEVLKTDSDAQIPAINQOLEIN-----KNSAQIILKLERONIS 227
Db 1246 STDSASTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSL 1304
QY 228 YEAVLTNAGEYIKASSEAGIKLGOALQSIIVDAGDQQAVALQAOONNSPDNIATKELID 287
Db 1305 -ESISTSVSDSTASTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1357
QY 288 AAEFKVNELEKQEH-----TGLTDEPLVKKAEQISQAOKDIOEIKPGSGDIPVGPSSGA 342
Db 1358 SASSTSTSDERASTSTSLSGSTSTSL-----SDSTSTSTSDASTSTSTSTSTSTST 1409
QY 343 ASASSAAGALK-----SSNNGRISLLDDVDNEMAMALQGRFMSMEQNVNPNPAKAKLQ 399
Db 1410 SLSSGLSTSVSDSTSTSTSDASASTSESDERASTSTSLSGSTSTSTSTSTSTSTSTST 1454
QY 400 AMEQLTAMSDQLVGADGELPAEIOAIKDALQALQKPSADGLATAMQOVAFAAKYV-G 458
Db 1455 ISDSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1512
QY 459 GSAGTAGVQNNKQVLTAFSTSSSSSYAAALSDGYSAYVTLNLSYSESSGVSQASISQ 518
Db 1513 DSASTSSSESVSTSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1569
QY 519 TANPALSRSVRSGLIESQGRSADASQRAAETIVRDSQFLGDVYSRLQVLDLSMTIVENP 578
Db 1570 STDSASTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1620
QY 579 QANOEIMOKLTASISKAPQPGYPAYONSADLOKFAQLREFYDGERSLAESQENAFR 638
Db 1621 NSASTSTSLSESTSTSLSDSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 1672
QY 639 KQPAFIQOVLYNIASLSESGYLS 660
Db 1673 ESTSTSTSTSVASNSMTSTSL 1694
```

RESULT 5
US-09-308-375-2
; Sequence 2, Application US/09308375
; Patent No. 6300117
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Proteases from Gram-Positive Organisms
; FILE REFERENCE: GC394-PC1
; CURRENT APPLICATION NUMBER: US/09/308, 375
; PRIORITY FILING DATE: 1999-05-14
; EARLIER APPLICATION NUMBER: EP9719636.4

;; EARLIER FILING DATE: 1997-09-15
;; NUMBER OF SEQ ID NOS: 3
;; SOFTWARE: FASTSEQ for Windows Version 3.0
;; SEQ ID NO 2
;; LENGTH: 2285
;; TYPE: PR1
;; ORGANISM: Bacillus subtilis
US-09-308-375-2

Query Match 7.38; Score 235; DB 4; Length 2285;
Best Local Similarity 19.38; Pred. No. 1.6e-07;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

```
QY 10 SGPSVSSNOGSMPII--NGOIASNSE-----TKESTKASEASPSASSVSSWFLS 61
Db 56 SAIDTVQKRLKSYQVYKSTVTKMDGSEKLTQOKKNGEYLQRETKLIIN----- 109
QY 62 AKNALISLRDAIILKNSSPTDSLQLEASTSTSTVTTRAADYDEAKSNFDTAKSGLEN 115
Db 110 -RNTAKQEQEVKRLQATEKELQOVQKKTQVRLNQQQPTVYQKNRHCFFDIIYTTDPK 168
QY 116 SGLENAKTLAEYETKADLMALQDMERL-----ANSDPSNNHTEEVN 158
Db 169 TNSTSKTTTNYDQORRAIEQLKODLEKLRQGGIVTDTTISLARKIINTAQAQOIEALQ 228
QY 159 NIKKALEKQKPTIDKLNKL-----VTLOONKSLTEVLKTTDS--ADQIPAI 203
Db 229 NRIIRLDKSAVAKKNLKTTEIYELQROAQVNVQNLTRYGSSNGSSRQAOVDYLVAV 288
QY 204 NSOLEINKNSADQIILKLERONISYEAVLTNAGEYIKASSEAGIKLGOALQSI--VDAG 260
Db 289 NS-LNVSTGCS--NNIRSQISLMQFRELASNAQTAANGASFQALQYQTFMSNTYLSG 346
QY 261 D-----OSQAVALQAOONNSPDNIATKELIDAAETVYVNELEKQEHGLTDSPL 308
Db 347 SLFYGAISGLKEMVSOAIEIDTLMTN-----IRRVNEDPYKYNELQRESIDLCGT-L 398
QY 309 VKKA-----EEOISQAOKDIOEIKPGSGDIPVGPSSGASASAGAL 352
Db 399 SNKTTDIIQMTGDFGRMGFDESELSLTAKTQAVIQ--NVSIDLPTDVTTLTAAMLNF 454
QY 353 K-SSNNGRISLLDDVDNEMAMALQ-----GFRSMI----- 384
Db 455 NIANDSISIDKLNQVNNVAVTTLDANSIRKAGSTAFEGVLENDLIGYTTAIAST 514
QY 385 -EQPNV-----NNPATAKELQAMEADLTAMSDQLVGADGELPAEIOAIKDAL 430
Db 515 RESGNIVGNSLKTTFARIGNNOSSIKALEQIGISVKTAGCEAKSA-SDLISEVAGKMDTL 573
QY 431 AQALIKPSADGLA-----TAM--QOVAFAAKYVGGSGAGT-----AGVQKNV 471
Db 574 SDAQOQNTSIVAGIYQLSRFNAMNNFSLAONNAKTAANSTGSMSEQOKYADSLQARV 633
QY 472 KQLYK--TAFSTSSSYAAALSDGYSAY-VTLNLSYSESSGVSQASISQ 520
Db 634 NKIQNNFTFEFAIAMSDF---ISDGLIEFTQAAGSLNASTGVIKSVGFPLPLAAVSTA 690
QY 521 NPALSRSVRSGLIESQGRSADASQRAAETIVRDSQFLGDVYSRLQVLDLSMTIVENP 578
Db 691 TLLSKNTRTLASSLILCTGRAMGOETLATAGLEGMTFAAASVRLKALGLLVTSTLVG 750
QY 559 DVSYRL-QVLDLSMTSTIVSNPOANOEIMOKLTASISKAPQPGYPAYONSADLOKFAAQ 617
Db 751 GAFALGWALESLLISSFEAKKAKDD-----FEOSQQTNVEALITTKNDSTDKLIQ 801
QY 618 L-REFYDGERSLAESQENAFRQPAFIQOVLYNIASLSESGYLS 658
Db 802 YKELQKVESRSLNSDEQEVLTQ---VTQQLAQGFEPALVKCY 840
```

RESULT 6
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C

Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 5080
LENGTH: 3696
TYPE: PRN
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

Query Match 6.8%; Score 219; DB 4; Length 3696;
Best Local Similarity 20.3%; Pred. No. 4,1e-06;

Matches 142; Conservative 125; Mismatches 307; Indels 126; Gaps 28;

QY 3 HHHHHHESPPES-----VSSNOSMNPILINQIASNSETKESTASERSPASSSVSS 55
DB 1084 HTNVNVQKRAQKQALIAKTNEKQSAI NSDNEETEEKOKAIQS--LNDAKNLADDOIYQ 1141
QY 56 WFLSSAKNAL-----IS-----LRDAIILK-----NSSP-----TDS 83
DB 1142 AASNVQVDAALNIGISINIKIQTNTFKQOARQOVNOKFOEKEAEALNSPHTARQDEKQDA 1201
QY 84 LSOLEASTSTSTVTVRYAANDYDEAKS--NFDTA-KSGLE-----NAKTLAEYETKMD 133
DB 1202 LVRITQAKET-----ALNDINQAOQNVVDVLTALSGIONIQNTQYVNRKKQAKTTIND 1255
QY 134 LMAAODMERLANSPPSNHTEEV--NIIKALEAQKDTIDKLKLVTLQNOKSLTEVVK 192
DB 1256 IYQ--OHKOSIONNDATTEEEKQEVANNLVNA--SQONVISKIDN--ATTNNQIDGI----- 1305
QY 193 TTDSADQIPAINSOLEINKNSADQIKLEPRONISYEAVLTNAGEYIKASSEGIKIGQA 252
DB 1306 VSDGRQSIINATIPDTISIKRNAKNDIDIKAKDKIKIQRINDATDELIQ--EANKRIEER 1362
QY 253 LOSIVDAGDQSOAAVLQAOONNSPDNIAATKELIDAEETKVELKQEHGTGLDPSLVKKA 312
DB 1363 K---IEAKDNIOKNSYRDQVNEAKTNGINKIENTIPATTIVKSEARQ-----AVQNNKA 1411
QY 313 EEOISQAQADIQEIKRPSGSDIPIVPGSSASASAGSAGACLKSSNNSGRISLLIDVDNEM 372
DB 1412 NQOINH---IQNTPDATNEEKQEA INRVSAELARVQAOIMEHTQOGVKTIKDDAITSL 1467
QY 373 AAIALOGFSMIEOFNVNVPATAKELQAMEAOQLTAMSDQLVGDELPAETQAIKALQ 432
DB 1468 SRINAO-----VEKESARN-----AIEQKTIQOT--QFINNDVATTEEEKVANNLVI 1514
QY 433 ALKOPSADSLATAMGOVAAPAAKVGSGSAGTAGTVOMNVKOLYKTAFSSSTSSSYAALS 492
DB 1515 ATRKOKSLDININSSNNVDENKAVG-----INEIANVLPATAVKSKAKKDIDQKLA 1566
QY 493 DQYSATKTLINSLXSESRGVQSAISOTANPALSRVSRSRGIESQGSASASQRAAETIYR 552
DB 1567 QOINOIQOTHTTTEKEAIAIOLANOKSNEA-----RTAIONEHSNNGVAAQAKSNGIHE 1620
QY 553 DSGTGLDGVSRLOVNDLSL-----MSTIVSNQAOAMEELIMOKLTASISKAPQFGYPAYO 605
DB 1621 IELVMPDAHAKKSDAKOSIDNKINKNEOSNTINTTTPDATDEE--KQALDKLTIKADAGYNNKV 1679
QY 606 NSADSLQKFAAOLEREFEVDGERSLASEQENAEFRKQAPFIQ 645
DB 1680 QAQTNQO--VSDAKTEAID--TITNIOANVAKKPSARVE 1714

RESULT 7

US-08-687-956A-23
Sequence 23, Application US/08687/956A
Patent No. 5861157

GENERAL INFORMATION:
APPLICANT: BURNIE, JAMES P
APPLICANT: MATTHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
STREET: FLOOR
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689, 6
FILING DATE: 28-JAN-1994

ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773

REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000

TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:
LENGTH: 1566 amino acids
TYPE: amino acid

STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal

ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-23

Query Match 6.1%; Score 194.5; DB 2; Length 1566;
Best Local Similarity 21.0%; Pred. No. 5.8e-05;
Matches 157; Conservative 116; Mismatches 281; Indels 195; Gaps 34;

QY 13 ESYVSSNOSMNPILINQIASNSETKESTASERSPASSSVSS 58
DB 40 ESTSTSGVNTAVVGTETGNPATNLDPKODNPSQAQFSGAQAGKGTGAMVDVSTSEL 99
QY 59 LSAKNA-----LISLADALINKNSPT--DSLQLEASTSTSTVTVRA-----AKQDVE-- 106
DB 100 DEAKSAQAGVIVSODAVVDGTVETSDKANQKETEIRIDYSKQAAIDIOKTTEDYKAAV 159
QY 107 AKSNFPAKSGLENATKTLAEYETKMDLMAALQDMERLANSPPSNHTEEV----- 157
DB 160 ARNQATETDITTOENAKKQAYE-----QDLANKRAEYERTITNENAKRADYEAUKLAQYOKDL 216
QY 158 -----NIIKALEAQKDTIDKLKLVTLQNON-----KSLTEVL--KTDSADQIPAINS 205
DB 217 AAVQANNDQAAVYAAKBAVDK--ELARVQANAANAARKEYEBAALANTTKNQIKAENA 274
QY 206 QLEINKNSADQIIKLERONISYEAVLTNAGEYIKASSEAGIKIGALQSIYDAGQOQA 265
DB 275 AIO-QRNA--QAKADYEAUKLAQYEKDL-----AAOSGNATNEDADVQAKKAAYEDEL 323

Db 34 OKVFADETTTSDVDTKVVGTOTGNPATNLPDAOGSAS-----KQAEOSQTKLERQ 84
QY 73 ILNKSSPTDLSLEASTSTSTVTRVAKODYDEAKSNFDTAKSGLENAKTLAEVETKMA 132
Db 85 MHTIEVTKTDDQAKAKSAGVNV---QDADVKKGTVKTAEEAVQETELKEEDYTKQA 142
QY 133 DLMAALQMERLANSPPSN--NHTEEVNNIKKALEAQKDTIDKLKLTVLTONKSTLEV 190
Db 143 -----EDIKTTDOYKSDVAHAHEAVAKIKAKNATKEQYCK--DMA-----HKAEVER 190
QY 191 LKTTSDADQIPAINSOLEINKNSADQIKLERONISYEAVLTNAGEVIKASSGAIKLG 250
Db 191 INANAASK--TAYEAKLAQYADLAAYOKTNAANQASQKALAAVQAEIKRVQEAANA 249
QY 251 QALOSIVAGOSQAAVYQAOQNNSPDNIATKELIDAEIKRVNLEKDEHGTIDSPLYK 310
Db 250 AAYDTAVANNAKNTETIAANAEIRKRATKAEL---ETKLAQYAEIKRVQOANAN 305
QY 311 KAEQD--ISOAKDIOEIKPSGSDIPYVPGSGSASAGALKSSNNSGRISILLDV 368
Db 306 EADYQAKLTAYOTELARQKANAD-----AKAAYEAANANNA----- 346
QY 369 DNEAALALOGFRSMIEGFVNNNPATAK-ELQAMEAQLTANSDDLVCADGELPAETQAIK 427
Db 347 -----ALTAEHTAIKQRENAKATYEALQYEDADLAAYKANAANEADYQAKLTAYQ 399
QY 428 DALQALQKPSADGLATAMGOVAFRAAKVGGSGAGTACTVQ-----MNVKOLY--KT 477
Db 400 TELARVOK-----ANADAKAAVEAAYANNAANALTAENTAIKKRNAAKADYEAKL 452
QY 478 AFSTSSSYAALSDGYSAVKTLSLSESRGVSQASISOTANPALSRVSRSRGIEQ 537
Db 453 AKYQADLAKYOKDLAD---YPVKIKAYEDEDASTIKALAE-----LKHKNEDC----- 498
QY 538 RSADASQRAAETIYRD-----SQTLDVYSRLQVLDLSLSTIVSNPOANOEIMQ 587
Db 499 ---NLTESQNLVYDLEPNANLSLTGDKFLKASAVODARSKTSKAKYDQ-KILQ 551

RESULT 14
US-08-466-390-4
: Sequence 4, Application US/08466390
: Patent No. 5686562
: GENERAL INFORMATION:
: APPLICANT: TOUTKATLY, GARY
: APPLICANT: LIDGARD, GRAHAM P
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/466,390
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESO, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MFP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

Query Match 5.7%: Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

Db 29 QIASSEETKSTKSEASPSASSSVSSWFLSSA---KNALISRDALINKNSPTSLTS 85
Db 338 QDALNELTEHRSKATQEWLEKQAOLE--KELSAALQDKCKLEKNELQGLSLQLEHLS 395
QY 86 QL-----EASTSTVTRVAKYDEAKSNFDTAKSGLENAK 122
Db 396 QLODNPQEKGEVLGDVLETLKQEAATLANNTOL-----QARVEMLETERGOOEAK 449
QY 123 TLAE---YETKMAIDLMALODME--RLANSPPSNHTEEVNNIKKA-LEAO---KDTIDK 173
Db 450 LLAERGHFEERKQOLSLITDQSSISMLQAKELEBOASOHGRLTAQVASLTSLETT 509
QY 174 LNKLTVLTONQ-----NKSLEVLKTTDSAD-----QIPAINSOLEINKNSADQ 216
Db 510 LNAITQQDDQELAGLKQQAKEKQAOQLAQTLQOQEOASGLRHQVQOLSSLSKQKQOLKE 569
QY 217 IKDLERONISYEAVLTNAGEVIKAS--SEAGIK-----LQALOSTVD 258
Db 570 VAKEQATROHQAQOLATAEEREASLERDALQKOLEKEKAKELEILQOOLVANE 629
QY 259 AGDOSQAAVYQAOQNNSPDNIATKELIDAEIKRVNLEKDEHGTIDSPLYKKAEOIS- 317
Db 630 ARDSAQSVTQAOEKA-ELSRKVEILOACVETARQOEHEAOVAAELEIQLRSEQKAT 688
QY 318 -----QAKDIOEIKPSGSDIPYVPGSGSASAGS-----AAGALKS-----S 355
Db 689 EKERYAQEKDQLOEOLAKES-----LKYTGSLSEKKRRRAADLEEQOQCIS 737
QY 356 NSGRISILLD-----DYDNEAALALOGFRSMIEGFVNNNPATAKEL-----QAMEA 403
Db 738 ELKAETRSLYBOHRRERKELEERA--GRKGLFARLLQLGBAHQAEFVLRRLEAEANAA 795
QY 404 QLTAMS--DQVAGADGELPAETQAIKDALQALQPSADG-----LAT-----ANGQ 448
Db 796 QHTAESECEOLY-----KEVAAMRDQYEDSQOEEAQYGMFOEQLMTLKECEKAQOE 848
QY 449 VAFPAAKVGG-----GSAGTACTYOMNVKOLYKTAFFSTSSSY 487
Db 849 LOEAKERYAGIESSELDISQOQKLAELHNLALRALQOQVEKEVRACKLADLSTLOEK 908
QY 488 AALSDGYSAVKTLSLSESRGVSQASISOTANPALSRVSRSRG-----IESQGRSA 540
Db 909 MAATSKYVARLETL-----VRKAGEQ--QETASRELVKEPPARAGDRQPEWLEQOGRQF 960
QY 541 DASQRAETIYRDSOTLDVYSRLQVLDLSLST--IYSNPANOEIMQKLTASISRAPO 598
Db 961 CSTQAAILOAMEREADEOMONTELERLA--ALMESQGOOQOEEQOEREVARLTQERGRA-- 1016
QY 599 FGYPVQVNSASLQKFAOLFREFYVDGERSLAEOENAFRQF--AFIQOVLVN 650
Db 1017 -----QADLALKEKARAELEM-----RLQNNALNQRVREFATLQDALAH 1054

RESULT 15
US-08-470-950-4
: Sequence 4, Application US/08470950
: Patent No. 5696439
: GENERAL INFORMATION:
: APPLICANT: TOUTKATLY, GARY
: APPLICANT: LIDGARD, GRAHAM P
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX


```

: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/470,950
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-470-950-4

```

```

Query Match      5.7%; Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

```

```

QY 29 QIASNETKSTKASBPASPSASSVSWFSLSA---KNALISLRDAIILNKNSPPTSLS 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 QDALNLTETBHSKATQEMLEKQKQLE--KELSAALQDKCLEKNEIKLSQLEHHS 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 QL-----EASTSTSTVTVAAKDYDEAKSNFTASGLENAK 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 QLODNPPEKEGVLGVLETLKQEAATLAANNLQ-----QARVEMLETERGOOEAK 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 TLAE---YETKMDLMAALQDME--RLANSDPSNNHTEEVNIIKA--LEAO---KPTIDK 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 450 LLAERGHFEERKQOSSLITDLOSSISNLSQAKEELEQASQAAGARLTAOVASTSELT 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 174 LNKLVTLQND-----NKSLEVLKTTDSAD---QIPAINSOLEINKNSADQ 216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 510 LNAITIQOQDELQGLKQAKQKQALQTLQOQEQASQGLRHQVEQLSSLSKQKEQDLKE 569
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 217 IITDLERQNSTYFAVLNAGEVIKAS---SEAGIK-----LGAALQSIYD 258
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 570 VAKQEAETRODHQAQALATAEEEREASLRERDAALKQLEALEKEKAKLETLQOOLQVANE 629
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 AGDQSAALVLAQOONSPDNIATKELIDAETKVNELKQEHGTGTPSVYKKAEBQIS- 317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 630 ARBSAOTSVTQAQREKA--ELSKRVEBELQACVETARQDQEHQAQVAVLEQLQTSSEQKAT 688
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 318 -----QAQKDIOETKPSGSDIPYVPGSSAASAGS-----AAGALKS-----S 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 689 EKEBVAQOEKQLOEQLOALKES-----LKVTKGSLSEBKRRADALDEQOCIS 737
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 356 MNGGRISLID-----DVNDEMAIALOGFSMTIEQFVNVNPTAKEL-----QAMEA 403
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 738 ELKAEETRLIVEQHKRREKKELEERA--GRKGLARLQJLEAHQAETEVRLRELAEMAA 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 404 QLTAMS--DQVLGADGLPRAEQIAIKDALAQALKOPSADG-----LAT-----AMGQ 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 796 QHTAESECEBLV-----KEVAAWRDGYEDSQEEAQYQAMQOEOLMTLKECEKARQOE 848
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 449 VAFPAAKVGG-----GSACTAGTVQNMVKQLYKTAFTSSTSSSY 487

```

```

DB 849 LQAKKQVAGIGSHSELQISROONKLAELHANLARALQOQKEVRAOKLADLSTLOEK 908
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 488 AALSDGYAVYTKRLNLSYESRSGVOSALSOFTANPLSSVSRSG-----IESGRSA 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 909 MAITSKEVARLETL-----YRKAGEDQ---QETASRELYKEPARAGDRQEMLEBQGRQF 960
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 541 DASORAETIVRDSQTLGDVYSRLQVLDLSMST--IVSNPQANQOEIMOKLTASISKAPQ 598
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 961 CSTQALQAMERAREAQMGNELERLRA--ALMESQGGQOQEEBQOQREVARLQJERGERA-- 1016
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 599 FCTPAVQNSADSLQKFAQLEREFTVDGERSLAESEQENARKQP---AFTQOVLVN 650
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1017 -----QADLALKEKARAELEM-----RLQNALNEQREVERPATLQELAH 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 16

```

US-08-467-781-4
: Sequence 4, Application US/08467781
: Patent No. 5780596
: GENERAL INFORMATION:
: APPLICANT: TONKATLY, GARY
: APPLICANT: LIDGARD, GRAHAM P
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
: STREET: 125 HIGH STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02110

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/467,781
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 424

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESQ, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100

```

```

: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein

```

```

US-08-467-781-4

```

```

Query Match      5.7%; Score 183.5; DB 1; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

```

```

QY 29 QIASNETKSTKASBPASPSASSVSWFSLSA---KNALISLRDAIILNKNSPPTSLS 85
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 338 QDALNLTETBHSKATQEMLEKQKQLE--KELSAALQDKCLEKNEIKLSQLEHHS 395
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 86 QL-----EASTSTSTVTVAAKDYDEAKSNFTASGLENAK 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 396 QLODNPPEKEGVLGVLETLKQEAATLAANNLQ-----QARVEMLETERGOOEAK 449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 123 TLAE---YETKMDLMAALQDME--RLANSDPSNNHTEEVNIIKA--LEAO---KPTIDK 173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 450 LLAERGHFEERKQOSSLITDLOSSISNLSQAKEELEQASQAAGARLTAOVASTSELT 509
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```



```

GENERAL INFORMATION:
APPLICANT: TOURKATLY, GARY
APPLICANT: LIDGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, ESQ. EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-924-4

Query Match          5.7%; Score 183.5; DB 2; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

QY 29 QIASNSETKSTKASEASPSASSVSSWSFLSA---KNALISLRDAIILKNSSPTDIS 85
DB 338 QDALNLTEBHSKATQEWLEKQALE--KLSAALQDKCKLEKKNIILOGKLSQLEHLS 395
QY 86 QL-----EASTSTSTVTRVAAKDYDEAKSNDTAKSGLENAK 122
DB 396 QLODNPPEKEGYLDVLOLETLKQEAATLAANNIOL-----QARVEMLETERGQOEAK 449
QY 123 TLAE---YETKADLMAALQDM--RLANGDPSNNHTEEVNIIKKA--LEAO---KDTIDK 173
DB 450 LLAERGHFEKEKOQLSSITLDQSSISNLSQAKEELQASQAHGALTQAVASLTSELT 509
QY 174 LNKLVTLQNO-----NKSLEVLKTTDSAD-----QIPALNSOLEINKNSADQ 216
DB 510 LNAITQQOQOQELAGLKQAKKEKQAOQLAQTLQOQEQASQGLRHQVEQLSSLSKKEQOLKE 569
QY 217 IIKDLERQNISEYAVLTNGEVIKAS---SEAGIK-----LGOALOSIYD 258
DB 570 VAEKQATRODHQOQLATAAEERASLRERDAALKOLEALEKEKAAKLEILOQOQLVANE 629
QY 259 AGDSQAVALQAOQNNSPNIAATKELIDAETKVNELKOEHTGLTDSPLVYKAEBOIS- 317
DB 630 ARNSAOTSVTQAOOREKA--ELSKRVEELQACVETARQEHQAQVAVLELQTLRSBOQKAT 688
QY 318 -----QAQKDIQETIKPSGSDIPIVPGSGSASAGS-----AAGALKS-----S 355
DB 689 EKERVAQOEKQLOEQOLAKES-----LKYTKGSLKEEKRRADALEEQORCIS 737
QY 356 NNSGRSLILD-----QVDNEMAAILOGFRSMIEQFNVNPNPRAKEL-----QAMEA 403
DB 738 ELKAEITRSLIVEQHKRERKELEERRA--GRKGLEARILQIGEAHQAEETVLRRELAERMAA 795
```

```

QY 404 QUTAMS--DQYVAGDELAEIQAITKDALQALKOPSANG-----LAT-----AMQO 448
DB 796 QHTAESECEQLV-----KEVAAMRDGYEDSQOEQAQYGAHFQEOELMTLKECEKARE 848
QY 449 VAEFAAKVGG-----GSAGACTVQMMVVKOLYTAESTSSSY 487
DB 849 LOEAKKEKVGIESHSELQISROQNKLAELHANLARLQOYQEKYVAKOALADDLSLTQEK 908
QY 488 AALSDGYSAYKTLNLSYSESRGVQSAISQTPANPALSRSVSRG-----IESQGRSA 540
DB 909 MATSKVARLETL-----VRKAGEQ---QETASRELVEKPARAGDRQEWLEEQGRCQF 960
QY 541 DASQRAETITVDSQTLGDIYSRLQVLDLSMT--IVSNPQANQEBIMQKLTASISKAPQ 598
DB 961 CSTQAALQAMERAEQMGNELETLRA--ALMSQGOQOEERQOQEREVARLTQERGCA-- 1016
QY 599 FGYPVQVNSADSLQKPAQLEREFYVGERSLAESQENAPRKOP---AFIQOVLVN 650
DB 1017 -----QADLALAEKARALEM-----RLQNALNEQRYVERATIQEALAH 1054

RESULT 19
US-09-452-294-1
; Sequence 1, Application US/09452294
; Patent No. 6287790
; GENERAL INFORMATION:
; APPLICANT: Lelievre, Sophie
; APPLICANT: Bissell, Mina
; TITLE OF INVENTION: UTILIZATION OF NUCLEAR STRUCTURAL PROTEINS FOR TARGETED
; TITLE OF INVENTION: THERAPH AND DETECTION OF PROLIFERATIVE AND
; FILE REFERENCE: IB-1454- Sequence Submittal
; Patent No. 6287790
; CURRENT APPLICATION NUMBER: US/09/452,294
; CURRENT FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/110,420
; PRIOR FILING DATE: 1998-11-30
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-452-294-1

Query Match          5.7%; Score 183.5; DB 4; Length 2101;
Best Local Similarity 20.8%; Pred. No. 0.00051;
Matches 161; Conservative 121; Mismatches 282; Indels 211; Gaps 33;

QY 29 QIASNSETKSTKASEASPSASSVSSWSFLSA---KNALISLRDAIILKNSSPTDIS 85
DB 338 QDALNLTEBHSKATQEWLEKQALE--KLSAALQDKCKLEKKNIILOGKLSQLEHLS 395
QY 86 QL-----EASTSTSTVTRVAAKDYDEAKSNDTAKSGLENAK 122
DB 396 QLODNPPEKEGYLDVLOLETLKQEAATLAANNIOL-----QARVEMLETERGQOEAK 449
QY 123 TLAE---YETKADLMAALQDM--RLANGDPSNNHTEEVNIIKKA--LEAO---KDTIDK 173
DB 450 LLAERGHFEKEKOQLSSITLDQSSISNLSQAKEELQASQAHGALTQAVASLTSELT 509
QY 174 LNKLVTLQNO-----NKSLEVLKTTDSAD-----QIPALNSOLEINKNSADQ 216
DB 510 LNAITQQOQOQELAGLKQAKKEKQAOQLAQTLQOQEQASQGLRHQVEQLSSLSKKEQOLKE 569
QY 217 IIKDLERQNISEYAVLTNGEVIKAS---SEAGIK-----LGOALOSIYD 258
DB 570 VAEKQATRODHQOQLATAAEERASLRERDAALKOLEALEKEKAAKLEILOQOQLVANE 629
QY 174 LNKLVTLQNO-----NKSLEVLKTTDSAD-----QIPALNSOLEINKNSADQ 216
DB 510 LNAITQQOQOQELAGLKQAKKEKQAOQLAQTLQOQEQASQGLRHQVEQLSSLSKKEQOLKE 569
QY 217 IIKDLERQNISEYAVLTNGEVIKAS---SEAGIK-----LGOALOSIYD 258
DB 570 VAEKQATRODHQOQLATAAEERASLRERDAALKOLEALEKEKAAKLEILOQOQLVANE 629
QY 259 AGDSQAVALQAOQNNSPNIAATKELIDAETKVNELKOEHTGLTDSPLVYKAEBOIS- 317
DB 630 ARNSAOTSVTQAOOREKA--ELSKRVEELQACVETARQEHQAQVAVLELQTLRSBOQKAT 688
```



```

; APPLICATION NUMBER: US/08/006,676B
; FILING DATE: 15-JAN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Oster, Jeffrey B.
; REGISTRATION NUMBER: 32,585
; REFERENCE/DOCKET NUMBER: REED-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 232 7845
; TELEFAX: (206) 236 0205
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-006-676B-1

```

```

Query Match      5.7%; Score 182.5; DB 1; Length 955;
Best Local Similarity 21.5%; Pred. No. 0.00019;
Matches 158; Conservative 122; Mismatches 269; Indels 187; Gaps 36;

```

```

QY 10 SGPESVSSQSSMNPITINGQIASNSETKSTK-----ASEASPSASSVS 54
DB 292 AGSERVAQSQ-----VEGQ-----QFKETHINLSLTTLGRVIDYLDAMATKGAQAQYS 340
QY 55 SMSFLSSAKNALISLSDALINKNS-----SPT-----DSISQLEASTSTSTYTRVA- 100
DB 341 VAFPRDSKLTFT--LKDS--LGNSKTFMTATVSPSALNTEETLSTLRVSRARDIVNVAQ 397
QY 101 -----AKDYDEAKSNFTPAKSGLENA-----KTLAEYETKMAIDMAALQOMERLA 145
DB 398 VNEDPRARIRLELEQMEDMQMAGDPAYVSELKKLALLESEOKRAADQALER-- 455
QY 146 NSDPSNNHTEEVANNIKKALEAKOTIDKLNKLVTLQONOKSLTEVYKTTDSADQIPAINS 205
DB 456 ----EREHNOVERLIRATEAEKSELE--SRAALQOE-----EMTATRRQADKQMALNL 503
QY 206 QL-EINKNSADQIIKDLERONISYEAVLTNAGEVIKASSEAGIKQALQASIVDADQSO 264
DB 504 RLKEBPARKERELIKMAKKDAALSKVRRKDAEIASERE--KLESTVAQLERQERERE 560
QY 265 AAV--LQAOQN-----NSPDNIATKELIDAEETKVNELKQEHGTLSPLVKKAEQOI 316
DB 561 VALDALQTHQRKQLEALESSERTAAER--DOLLQQLTELQSERFQL--SQVVTDTRE-- 612
QY 317 SQAKNDIQIK-----PGSDPIPIYPGSGSASASAGALKSSNNGRISILLDDVON 370
DB 613 -RLTRQLORIYEGETELARVDALCAQOEMERHYAAVFLQTL-----LLELATWED- 665
QY 371 ENAATIALQCFRSMIEQFNVNNPATAKEIQAMEA-----OLTAMSDQLVGAD-- 416
DB 666 -----ALRE-RALAEER-----DEAAAELDAASTSONARESCERLTLSLEQUREEERA 715
QY 417 GELPAEIQAIKDA-----LAQALKQPSADGLATAMGOVAPPAKAVG--GS 460
DB 716 AELASQLEETAAKSAEODRENTATLEQOLRESEARAELASQLEATAPAAKMSAEODR 775
QY 461 AGAGTGVQNNVQLYKTA--FGSTSSSSYAALSS--DQSAVKTLNLSLSESR----- 509
DB 776 ENTRATLEQDLRDSERAEELASQLESTTAAKASABODESTRATLEQOLRDSERAEEL 835
QY 510 -SGVOSAIQOTANPALSRVSRSRGIESQGRSADASORAETTVIRDSQTLGDVYSRLQVLD 568
DB 836 ASQLESTTAAKMSAEODRESTRATLEQOLR--ESERAAE-----LAS 876
QY 569 SLMTSTVSNPQANQEE-----IMQKLTASISKAPQGYPAVONSADSLQKFAQLEREF 622
DB 877 QLESTTAAKMSAEODRESTRATLEQOLRDSERAEEL--ASQLEATAPAAKMSAEODREN 933
QY 623 VDG--ERSLAEQENA 636
DB 934 TRAALEQOLRDSERAE 949

```

```

RESULT 22
US-08-282-845-2
; Sequence 2, Application US/08282845
; Patent No. 5719263
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: A 230kD Antigen Present in Leishmania
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Immunex Corporation
; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Apple Macintosh Operating System 7.1
; SOFTWARE: Microsoft Word for Macintosh 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/282,845
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/006,676
; FILING DATE: JANUARY 15, 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 5004-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-282-845-2

```

```

Query Match      5.7%; Score 182.5; DB 1; Length 955;
Best Local Similarity 21.5%; Pred. No. 0.00019;
Matches 158; Conservative 122; Mismatches 269; Indels 187; Gaps 36;

```

```

QY 10 SGPESVSSQSSMNPITINGQIASNSETKSTK-----ASEASPSASSVS 54
DB 292 AGSERVAQSQ-----VEGQ-----QFKETHINLSLTTLGRVIDYLDAMATKGAQAQYS 340
QY 55 SMSFLSSAKNALISLSDALINKNS-----SPT-----DSISQLEASTSTSTYTRVA- 100
DB 341 VAFPRDSKLTFT--LKDS--LGNSKTFMTATVSPSALNTEETLSTLRVSRARDIVNVAQ 397
QY 101 -----AKDYDEAKSNFTPAKSGLENA-----KTLAEYETKMAIDMAALQOMERLA 145
DB 398 VNEDPRARIRLELEQMEDMQMAGDPAYVSELKKLALLESEOKRAADQALER-- 455
QY 146 NSDPSNNHTEEVANNIKKALEAKOTIDKLNKLVTLQONOKSLTEVYKTTDSADQIPAINS 205
DB 456 ----EREHNOVERLIRATEAEKSELE--SRAALQOE-----EMTATRRQADKQMALNL 503
QY 206 QL-EINKNSADQIIKDLERONISYEAVLTNAGEVIKASSEAGIKQALQASIVDADQSO 264
DB 504 RLKEBPARKERELIKMAKKDAALSKVRRKDAEIASERE--KLESTVAQLERQERERE 560
QY 265 AAV--LQAOQN-----NSPDNIATKELIDAEETKVNELKQEHGTLSPLVKKAEQOI 316
DB 561 VALDALQTHQRKQLEALESSERTAAER--DOLLQQLTELQSERFQL--SQVVTDTRE-- 612

```

```

RESULT 23
PCT-US94-00324-1
: Sequence 1, Application PC/TUS9400324
GENERAL INFORMATION:
APPLICANT: Reed, Steven
TITLE OF INVENTION: Diagnosis of leishmaniasis
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple System 7.1
SOFTWARE: Microsoft Word, version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00324
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/006,676
FILING DATE: 15-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 5004-WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 955 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-00324-1
Query Match 5.7%; Score 182.5; DB 5; Length 955;
Best Local Similarity 21.5%; Pctd. No. 0.00019;

```

Matches	158;	Conservative	122;	Mismatches	269;	Indels	187;	Gaps	36;
QY	10	SGPESVSSNOSSMPDINGQIATNSNETESTK-----	ASASPSSASSVS	54					
Db	292	AGSEKVAASO-----VEGO-----QFKEATHINLSTTLTGIVIDLMDATMGAKAAYS		340					
QY	55	SMSFLSSAKNMLISRDALINKNS-----SPT-----DSLQLEASTSTSYTTRVA-	100						
Db	341	VAFPHDSKLTFP-----LKDS-LGGSNRTFMATVSPALNTEETLSTLRYASRRARDIVNVAQ	397						
QY	101	-----ANDYDEAKSNFDTAKSGLENA-----KTLAEYETKADIMAAJLQDMERLA	145						
Db	398	VNEDPRARIRIELEQOMEDMQAMMGCPAYSELKTKLALLESEQKRAADQALER--	455						
QY	146	NSDPSNNHTEEVNNIKKALEAQDITDKLNLVTLQONOKSLTEYLKTTSDAQIPAIN	205						
Db	456	-----EREHNOVERLLRRTATEAKSELE--SRAALQOE-----EMTATRRQDKQKQALNL	503						
QY	206	QL-EINKNSAQIITIDLERONISYAVNLNAGEVTKASSEAGIKKQALQISIVDAQDSQ	264						
Db	504	RIKEQOARKERELKEMAKKDAALSKVRRRKADEATASENE--KLESTVAQLERQORE	560						
QY	265	AAV--LQAOQN-----NSPDNIATKTELIDAEFTKVNLEKQEHGTGTFSPVYKKAEEQI	316						
Db	561	VALDALQTHQKRLQLEALSSERTAAER--DOLLOQLTEILOGSERTOL--SQVYTBRE---	612						
QY	317	SOAQKDIOEIK-----PSGSDPIYVGPSSAASAGSAAGALKSSNNGRISLLDDVDN	370						
Db	613	-RLTRDQRIQYEVGETELARDVALCAQMEBARYHAAVFHLQT-----LLELATWEMED-	665						
QY	371	EMAIALDQFSSMLEQFVNNPATAKELQAMEA-----QLTASDQLVQAD---	416						
Db	666	-----ALRE-RALAE-----DEAAAAELDAAASTSQNARESCERLTSLSEQLRSEERA	715						
QY	417	GELPAEIAQIDA-----LAQALKOPSADGLATAMQVAFPAAKVGG--GS	460						
Db	716	AEIQAQLEATPAAKSASADQDRENTATLTLEQULRESEBARAEIASOLEATNAKMSAEODR	775						
QY	461	AGTACTVOMNKKOLKTKTA--FSTSSSSVYAAALS--DGYSAKYTKLNSLYSES-----	509						
Db	776	ENTRATLTLEQULRDSERAAELASOLESTYPAKMSAEQDRESTRATLEQULRDSERAAEL	835						
QY	510	-SGVOSAIISQTPANPALSHSVSRGIESQGRSADASQRAAEITVRSQTLGDVYSLQVID	568						
Db	836	ASQLESTYPAKMSAEQDRESTRATLEQULR--ESEERAAE-----LAS	876						
QY	569	SLMSTIVSNPQANDQEE-----IMQKLTASISKAPQFQYPAQNQADSLOKFAAQLEREF	622						
Db	877	QLESTYPAKMSAEQDRESTRATLTLEQULRDSERAAEL--ASQLEATPAAKSSAEQDREN	933						
QY	623	VDG--ERSLAESQENA	636						
Db	934	TRALEQULRDSERAA	949						
RESULT	24	US-09-336-447A-5							
		Sequence 5, Application US/09336447A							
		Patent No. 6310190							
		GENERAL INFORMATION:							
		APPLICANT: HANSEN, ERIC J.							
		APPLICANT: AEBI, CHRISTOPH							
		APPLICANT: COPE, LESLIE D.							
		APPLICANT: MACIVER, ISOBEL							
		APPLICANT: FISKE, MICHAEL J.							
		APPLICANT: FREDENBURG, ROSS A.							
		TITLE OF INVENTION: USPA1 AND USPA2 ANTIGENS OF MORAXELLA CATARRHALIS							
		FILE REFERENCE: AMCY:024							
		CURRENT APPLICATION NUMBER: US/09/336,447A							
		CURRENT FILING DATE: 1999-06-21							
		NUMBER OF SEQ ID NOS: 98							
		SOFTWARE: PatentIn Ver. 2.1							
		SEQ ID NO 5							

```

; LENGTH: 892
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-5

Query Match          5.7%; Score 182; DB 4; Length 892;
Best Local Similarity 19.3%; Pred. No. 0.00019;
Matches 126; Conservative 110; Mismatches 240; Indels 178; Gaps 28;

OY 7 HMEGPEVSSNOSMNPIT-----NGOIASNSEFK-----STRASESPSASS 51
DB 271 HMTAGKATAVASAKVNGLTGNGFAGASKTGNGTVSGSENNRQIVNAGAGNISADSTD 330
OY 52 SVS-----SMSFLSSAKNALI-----SLDALINKSSPTDISLQLEAST 91
DB 331 AVNGSOLVALATAVKADADENEFALTKRONTLLEGEADADALIAON-----QTDIYA 382
OY 92 STSTVTRVAKDYDEAKSFND--TAKSGLENAKTLAEYETRMADLMALADM--ERLAN--- 146
DB 383 NKTAIR-----NFNRTVVGFELEKKNKAGIAKNQADITOTLENNVGEELLNLSG 431
OY 147 -----SDPSNN-----HTEEVNNIKKALEAKDTIDKLNKLVTLQONKSL 187
DB 432 RLDDQKADIDNNINNYDLAQODQSSDIKTKKNVE--EGLLDLSGRLI--DQKADL 486
OY 188 TEVLKT-----TDSADQIPAINSOLEINKNSD-----QIYDLERONISTEAVLT 233
DB 487 TDKIKLENNVEGLDLSGRL--IDQKADIAKNQADIAONQTDIDDLAAYNELQDYAQ 544
OY 234 NAGEV-----KASSEAGIKGALQISIVADGQSOAAVLOAOONNSPDNIATKELIDAA 289
DB 545 KQTEAIDALNKASSANTDRIATATLGT--AENKKQAIKAKOAKENKKGDIKAKQADITQLH 602
OY 290 ETKVNELKQEHGTL-----TDSPLYKAEQISOAKDIOEIKPS---GSDI 333
DB 603 DKKITNLGILHSVAVRANGNNTQGVATNKADIAKNQADIANINKIYELAQQODQSSDI 662
OY 334 PIVPGSGSAASAGSAGALKSSNNGRISILLDDVDNEMAALLOQFRSMIEFPVNNNA 393
DB 663 -----KTLAKVSAANTDRIAKNAKEADASFETLRKQNTLLEQG--EALVEQ---NMA 710
OY 394 TAKELOAMEAQLTAMSODLVGADGELPAETIOAIKDALAOLKOPASGLATAMGOVAFAA 453
DB 711 INDELGFPAHADVDQKOL-----QNDADITTKAAITEQINFTVANGF----- 755
OY 454 AKVGGSGAGTAGTQVQ-----NVKQLYKT--AFSSTSSSYAAALSDGYSAYKTINSL 504
DB 756 -EIEKKKAGIATKQELILQNDRLNOINETNNRQDKIDOLGYALK--EQGQHNNRISAV 813
OY 505 YSRSRGVOSALSOTANPALSR-----SVRSRGIESQGRS 539
DB 814 EROTAGGIANAIAIATLPSPSRAGEHHVLFGSGYHNGOAAVLSGAGLSDPTGS 867

RESULT 25
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-Hwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328, 254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match          5.6%; Score 180; DB 1; Length 2482;
Best Local Similarity 19.9%; Pred. No. 0.0011;
Matches 151; Conservative 125; Mismatches 288; Indels 196; Gaps 32;

OY 13 ESVSSNOSMNPILINGQISNSEFKESTKASEASPSASSSVSWFLSSAKNALISLRDA 72
DB 317 ETLTLEKKEMSSIT--LNKREIELTIOENGTLKEINNS-----LNQKNLNLQKSS 367
OY 73 ILKNSSPDLSQ-----LEASTSTVTRVAAKDY-----DE 106
DB 368 FANYIDERKKSISELDQYKOEKLLILQRCFEFGNAYEDLSQYKAAQENSKLECLINE 427
OY 107 AKSNEDTAKSGLEN-----AKTLAEYETKMA-----DLMAALOMERLANS--DPSN 151
DB 428 CTSLCENRKNELQLEKAFKEHOFELTKLAFABERNQMLLELTVQDALRSEMTDNON 487
OY 152 NHTEEVNNIKKALEAKQDTIDKLNKLVTL--LQONKSLTEVLKTTVSADQI--PAINS-- 205
DB 488 NKSSEAGIGLQELMTLKEQONKQKQEVNDLQENQELKVMKTKHQCQALSESPRINSVK 547
OY 206 -----OLEINKNSADQIITKDLERONISTEAVLTNAGEVITKASSEAGITKG 250
DB 548 ERESERNQCNFKPQMDLEVEKEISLD-----SYNAQLVLEAMLRNKLKLOESKEKECLO 603
OY 251 QALOSIVADGQSOAAVLOAOONNSPDNIATKEL--IDAAE----- 290
DB 604 HELQITRGDLETSNLODMQSOE-----ISGLKDCETDAEKKYISGPHELSTSONDAHL 657
OY 291 -----TKVNELK-----QEHGTGLTDS-----PLYKKAEOIISOAKDIOEI 326
DB 658 QCSLOTMMKNKINLEKICELQAEKELYVELTENDSSECIATRKAAEEVG--KLINLV 714
OY 327 KPSGSDIPIVQ-----PSGSAASAGS-----AGALKSSNNGRISILLDDVDNEMA 374
DB 715 KIILNDSGLLHGLVLDIGGEGEPQNECHPVSILAPRDESNSEYEHILTLDREKVOHNFAP 774
OY 375 -----IALQGRSMIEQFVNNPAPATAKELQ-----AMEAQLTAMSDQLVGADGELPAEIO 424
DB 775 LQKFLSLQSEHKTLHDQHCQSSKSEIQTVYDSILKAELVLTSLRNFOGGVLEKEMQI 834
OY 425 AIKDALAOLAKQ--PSADGLATAMGOVAFAA--KVGGS-----AGTAGTQVQNVKQ 473
DB 835 GLEEGLVPSLSSSCVYPDSSSL--SLGDSSEFYRALLEQOTDMSLSLLEAGVANSQCSYDE 893
OY 474 LYKTA--STSSSYAAALSDGYSAYKTLINLSYSSRSRGVOSALSOTANPALSRVSNS 531
DB 894 VFCSSLOEENLTKRTPSPAPAKVEELLESICEVYRQSLKLEK-- 937
OY 532 GIESQGRSADASORAETIVRSQTLGDVYSRLQVLDLSLMTSTIVSNPANOGEIMQKLT 591

```


Db 938 -MESGIMKNKEIJELEBOLLS-----SERQELDLCKROYLS-----ENFOWOQKLTLS 983
Oy 592 -SISAKPOFGYPAYONSADSLOKFAOLEREFEVD-GERSL 629
Db 984 VLEMEKSLAAEKKOTEOLESLEEVARLOGLDLSRSL 1023

RESULT 26

US-08-685-871-2
Sequence 2, Application US/08685871
Patent No. 6013499
GENERAL INFORMATION:
APPLICANT: NARUMIYA, Shuh
APPLICANT: IWAMATSU, Akihiko
TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685, 871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/845
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1354 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-685-871-2

Query Match 5.6%; Score 178.5; DB 3; Length 1354;
Best Local Similarity 19.8%; Pred. No. 0.0006; Indels 179; Gaps 34;
Matches 141; Conservative 123; Mismatches 269;

Oy 12 PESVSNOSNNPIINGOIASNSEKESTKASEASPSASSSVSSVSSAKNALISLRD 71
Db 384 PKAVYGNL--PEYGFYYGNRRRLSSANND--NRTSSNADKSLQESLOKTIYKLEE 437
Oy 72 AILKNSSPFDLSOLEASTSTSTVTRVAAKDYDE--AKSNFTAKSGLENAKTAEY- 127
Db 438 QLNHE-MOLKDEMO-KCRTSNIKIKDKI-MKELDEGQNRNRLLESTVSOIEKEMLLQHR 494
Oy 128 -----ETKADIMAAALODMERLA-NSDPSNNHTEEVNNIKKALEAQK 168
Db 495 INEQAKAEQENKRRNVEVSTLKQLEDLKKVSONSLAN--EKLSQLQOLEBAN 551
Oy 169 DTIDKLNLVTLQONKNSLFEVLKTTDSADQIPAINSOLE-----INKSADQIITKD----- 220

Db 552 DUL--RTESDPAVRLRKSHTEMSK---SISQLESINRELQERNRILNENSKSQTDKDYVL 606
Oy 221 -----LERONISY-----EAVLTNAGEVIRKASSEAGIKGALOSIVAGOSQAAV 267
Db 607 QALIEARRDRGHDSMIGDLQAKRTISLOEEVK-----HLKHNLEKV--EGEKREKAD 657
Oy 268 LOAOONNSPDNIA-----ATKELIDAEETKYNELKOEHTGLTD-----S 306
Db 658 MLNHSKKNLLEIDLANKKLKSLQGRLEGEVNEKHVTARLTKDQKSTIEAKSVAMCEME 717
Oy 307 PLVKAEFOISOAKODIOEIKPSGSDPIYVPGSGSASAGSAGALKSSNN-----SGRI 361
Db 718 KKLKEERARREKAEENRVQVQIEKQCSMLDY-----DLKOSQOKLEHLTKNK 762
Oy 362 SLULDVDVNEAAIYALQCFRSMIRQFNNNPATAKEL--QAMERQLTMSQVLGADDEL 419
Db 763 ERMEDEVKN-----LTLQ-----LEQSKRLLLONELKTOAFEA-----DNLGLEGKQ 807
Oy 420 PAETQAIKDA-----LAQALKOPADGLATMAGOVFAAAKVGSGSAGTAGTV----- 467
Db 808 KOEINTLLEAKRLEFELOALTKO-----YRGNEGQMRLED 844
Oy 468 QMNVKQLYKTAFFSTSSSSVAAALSDGYSAYKTLNLSYSESRS-GVQSAISOTANPALSR 526
Db 845 QLEAQYFSTLYKTOVKELKEIEEKNRENELKIOELQNEKETLATQDLAET--KAESE 902
Oy 527 SVSNGITSQGRSADASRAAEITVRSQTLGD---VYSRLQVDSLMTSTVSNQANQE 583
Db 903 QLARGLLEEQ--YEELQYESKKAASRNROETTDQHTVSRLEANSMTKIOEILRRENE 960
Oy 584 EIMOKLTASISKAPQFGYPAYONSADSLOKFAOLEREFEVGERGSLAESQEN 635
Db 961 ELTEM-----KKAEEYKL--EKEEISNLKAAFEKN-INTERTLKQAVN 1004

RESULT 27

US-09-310-187A-1
Sequence 1, Application US/09310187A
Patent No. 6358751
GENERAL INFORMATION:
APPLICANT: Benichou, Gilles
APPLICANT: Fedoseyeva, Eugenia
TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
FILE REFERENCE: UCSF-090
CURRENT APPLICATION NUMBER: US/09/310, 187A
CURRENT FILING DATE: 1999-05-12
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1939
TYPE: PRT
ORGANISM: Homo sapiens
US-09-310-187A-1

Query Match 5.5%; Score 175; DB 4; Length 1939;
Best Local Similarity 19.7%; Pred. No. 0.0017; Indels 142; Gaps 28;
Matches 130; Conservative 125; Mismatches 262;

Oy 35 ETKESTKASEASPSASSSVSSM-----SFLSSAKNALIS-LRD-----ATLN 75
Db 1352 ETEAKAEIQRLVLSKANSVYQMRKRYETDAQTORTBELEAKKKLQRLQDAEAEVAVNA 1411
Oy 76 KNSSPFDLSOLEASTSTSTV---TRVAAKDYDEANSNFTAKSGLENAKTAEYETKM 131
Db 1412 KCSSELEKTKHRLQNEIEDIADYVERSNAAALDKKORNF-----KILAEWKQY 1462
Oy 132 ADLMAALODMERLANSNHNTEEVNNIKKALEAQKDTIDKLKIVTLQONKNSLFEVL 191
Db 1463 EESQSELESSQKEARS-----LSTELFKLNAYE-----ESLHELEFKENKNLOE-- 1509
Oy 192 KTTDSADQIPAINSOLE-EIKNN--SADQIILDLERONISYEAULTNA-----GEYIK 240

```

Db 1510 -----EISDLTEQLGEGGKNVHELEKVRKOLEVEKLELQSLAEFAEASLEHEEGKIIR 1562
QY 241 ASSEAGIKRIGALQSLQIVDAGDOSQAALVLAQOONNS--PNNIATKELIDIAETKVNELKO 298
Db 1563 AOLEFNOIKAEIERKIAEEDEME---QAKRHQRYVDSLOTSLDAETRSRREIVLRVK 1618
QY 299 EHTG-LTDSPL-VKRAEQIISOAKDIOEIKPESGDIPIVPGSGAASAGSAGALKSSN 356
Db 1619 KKEGDLNEHIDQSHANRMAAEQKQVSLQSLKDTQI-----QDDAVR-AN 1666
QY 357 NSGRISLLDDVDNEMALALOGFRSMIOGFVNNPATKEL-----QAMEAQLTAMS 409
Db 1667 DDLKENIAIVERNNLLOAELELRAVVBQTERSRLAOLETSETSERVQLHSONTSFLI 1726
QY 410 DQ-----LYGADGELPAEIOAIKDALQALKQPSADGLATPMGVAPAAKVGSGAG 462
Db 1727 NOKKKNESLUTQLOSEVEEAVQECRAAEKAKK-----AITDAAMAEELKEQDT 1777
QY 463 TAGTQVM--NVKQLYKTAESSTSSSYAALSDGYSAYKTLNLSYSESRSGVQSAISOQA 520
Db 1778 SAHLEMRKKNMEQTIKD-LQHRLEAEQIALKGKKQLQKLEARVRELEGELEAEQKRMA 1836
QY 521 NPALSNSVRSQI-ESQGSADASQRAETIVRDSQTLGDVYSRLQVLDLSMSTIVSNFO 579
Db 1837 E-----SVKGRKSERIKELTYOTEE---DKRNL-----LRIDLVKDLQKVKAYK 1881
QY 580 ANOEIMOKLTASISKAPQGYPAVONSADSLQKFAQLEREFVDEGERSLAESQENAFR 638
Db 1882 RQAEFEAEQANNTLSK-----FRKVOHELDEAEERA-----DIAEQVAKLR 1923

```

RESULT 28

```

US-09-268-347-36
; Sequence 36, Application US/09268347
; Patent No. 6353182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRF
; ORGANISM: Haemophilus influenzae
US-09-268-347-36

```

```

Query Match          5.4%; Score 174; DB 4; Length 2411;
Best Local Similarity 19.3%; Pred. No. 0.0028;
Matches 143; Conservative 123; Mismatches 307; Indels 160; Gaps 30;

```

```

QY 1 MHHHHHMEGSPESVSSNOSSMNPINGOIAS-----NSETKESTRASEA 45
Db 562 LNRHVEDAYKGLNLEKNANKQPLVTDSTAATVGDRLKLGWVSTKNGKREESNQVKA 621
QY 46 SPBASSSVSMFLSSAKNALISL-----RDALINKNSPTDSLQLEASTSTSYT 97
Db 622 DEVLFTGAGAAVTYTSKSENGKHITVSAETKADSGLEKDG---DTIKLKVDMQNDNVL 678
QY 98 RYAAKDYDEAKSNFTDAKSGLENA-----KTLAETETKMA---DLMAALOD----- 140
Db 679 TYGNNGTAVTKGGFETVTKGADPDADRGKYTVKDATANDADKKYATYKDATATINSATVY 738
QY 141 -MERLANSPPSNHTEEVNNIKKALEAKQDITDKLNLVTLQONKSLTEVL-----KT 193
Db 739 KTEMLFTSIDENPTD--NGKDALKAGDPLTFKAKKNLKVKRDGNITFDLAKNIEVKI 796
QY 194 TDSADOI-----PAIN-----SOLEINKNSADQIIKLERONISYEAVLVN 234
Db 797 AKVSDTLTIGGNTPTGTTATPKVNTITSTADGINFAKETADA-----SGSKNVYLKGIATY 852

```

```

QY 235 AGEVIRKASSEAGIKRIGALQSLQIVDAGDOSQAAL-----VLOA-----QONNSPDNIATKELI 286
Db 853 LREPSAGAKSHVDLN-----VDATKKSNMAASIEDVILFAGNNIGONGNNVYVATYDTV 906
QY 287 DPAETKVNELKQEHGGLTDSPLYKKAEOISOAQ-----KDIOET--- 326
Db 907 NPTD-----DSGTITTVYVQKADGKADVKIGAKTSVIKDHNKLFETGDKLDANN 958
QY 327 -----KPSGSD-----IPIVGP-----GSAAGSAGSAGALKSSNSGRISLL 365
Db 959 GATVSEDDKDTGTGLVTAKYIVDAVYKSGMRYTGEGATAETGATRVANAGNAETVTSGIS 1018
QY 366 DDVDEMAIAL-----QGRSMIEGFVNNPNPATKELQAMEQLTAMSQDVLGADGELPA 421
Db 1019 VPKRKNATPTAYVSKDNGNINVKYDVNVGDGLKIGDDKIVADTTTLT--VYGGKVSVA 1076
QY 422 EIQAIDALQALKQPSADGLATPMGVAPAAKVGSGAGTAVQMMVVKQLYKTAES 481
Db 1077 GANSVNN---NKKLVNAEGLATALNNLSMTAKADKYADGESEGETDQEVKAGDKVTFKA 1132
QY 482 TSSSSYAAALSD-GYSAYKTLNLSYSESRSGVQSAISOTANPALSRVSRSGIE---SOG 537
Db 1133 GKNLKYKQSEKDEFTYSLQDTLGLTISITLGTANGRNDTGT-----VINKDGLITLILANG 1187
QY 538 RSA--DASQRAETIVRDSQTLG-----DVYSRLQVLDLSMSTI-VSNPOANOEIMOKL 589
Db 1188 AAGGTASNGNTISVTKDGISGKKEITVVKSKLTKYKDTQVTAGATQPAAMTAEVAKOD 1247
QY 590 TASISKAPQGYPAVONSADS 610
Db 1248 LVDLTK-PATG--MAGNGADA 1265

```

RESULT 29

```

US-08-938-105-3
; Sequence 3, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,105
; FILING DATE:

```

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Crook, Wanneil M.
; REGISTRATION NUMBER: 31,071
; REFERENCE/DOCKET NUMBER: 3595-4
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1886 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-938-105-3

```



```

OY 380 -----PFSMIEQFVNVPAT-----AELQAMEA 403
Db 298 EKALAKVSDNPOLDNSYTNKKLMNLNSLDTKVEEDNNTLSDDKKQALKOEIDTKOSIDR 3577
OY 404 QLTAMSDQVLGADGELPAEIOAIKDALQALKPESAGLGTATAMQVAFPAAKYVGSAGT 463
Db 358 ORNIITIDQNGASNNK-----OATEDILNSVPSKNEVEDIMKRI-----KTNGRS-N 403
OY 464 AGTYOMANVKOLKRAAFSTSTSSSTAALSDGYSAVKTLNLSISRSRGVSAISQTANPA 523
Db 404 EDIANQIAKQI--DGLATSTSSDDILTKMELDQSKESLIQOLLTRTGLGDEA-DRIAKKL 460
OY 524 LRSYSRSRGISQGRSADQORAETIYRDSQTDGVYSRLQVDSLSMT-----573
Db 461 LSONLSNSQIYEQLRHFNHSGGTATADDILNGVINDAKDKRQAIETILOTIRINKDKAKII 520
OY 574 --IYSNPQANEELIMOKLITASISAKPGYPAVONNSDSL---OKFAOLERE 621
Db 521 ADVIAIARVOKKSDIMDL-----HSAIEGRANDLIDIEKRAKQAKKD 562

```

RESULT 34
 US-08-460-309-4
 Sequence 4, Application US/08460309
 Patent No. 5837496
 GENERAL INFORMATION:
 APPLICANT: Engvall, Eva
 APPLICANT: Leivo, Ilmo
 TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
 TITLE OF INVENTION: Fragments and Uses Thereof
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Campbell and Flores
 STREET: 4370 La Jolla Village Drive, Suite 700
 CITY: San Diego
 STATE: California
 COUNTRY: USA
 ZIP: 92122
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,309
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/125,077
 FILING DATE: 22-SEP-1993
 APPLICATION NUMBER: US PCT/US 94/10730
 FILING DATE: 21-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/472,319
 FILING DATE: 30-JAN-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/919,951
 FILING DATE: 27-JUL-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Campbell, Cathryn A.
 REGISTRATION NUMBER: 31,815
 REFERENCE/DOCKET NUMBER: P-LA 9721
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 535-9001
 TELEFAX: (619) 535-8949
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3111 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 US-08-460-309-4

```

Query March 5.38; Score 170.5; DB 2: Length 3111;
Best Local Similarity 19.98; Pred. No. 0.0069;
Matches 129; Conservative 103; Mismatches 259; Indels 157; Gaps

QY 13 ESVSSNOSSMPDIINGOIASNEFEKKESTKSEKSPASSSVSSWFLSAAKNALISLDA 72
DB 1858 EDIQTKLPSEELNNDKIDIDLSOEINDRKIAEVSQAEHA-----A 1899
QY 73 ILNKNSTFDLSOLEASTSTSTVTYVAAKDYDEAKSNFDTKSGLENAKTLAEYETKMA 132
DB 1900 QLNSSAVLDGI--LDEAKNISFNATAAFAKYSINIKDYIDEAEKVAKEAKDLAHEATKLA 1957
QY 133 -----DLMALQ-----DMERLANSDPSSNHTTEVNNIKALE----- 165
DB 1958 TGPRLKLEKDGKGLCKSFRLNEARKLANDYKENE--DHLNGLTRIEDNARDNGDLLR 2015
QY 166 AOKPTIDKLN-----KLVTLONONKSLTEYLKTTDSADQIPAINSOLEINKNSADQI 217
DB 2016 TLNLTDLGKLSLIPNDTAAKIQNVKDKARQANDPAK--DVLADITELHQNLDGKKNYKYL 2073
QY 218 IKDLERONISTEAVLTNAGE-VIKASSEAGIK-LGQALQSIYDAGDQSOAAVLQAQONNS 275
DB 2074 ADSVAKTY---AVVKDPSKKKIIADADATVKNLEQADRLID-----KLKPIKELEDNL 2124
QY 276 PDNIATATRELDAAETVYNELKQEHGUDPSLVYKAAEQISOAQKDIOEIKPSSDIP 335
DB 2125 KKNISEIKELINOARKQANIKYSSVSGGCIRTYRPEIKKSYNNIYVNNKTAVAD--NL 2183
QY 336 VGPSSAASAGSAGALKSSNNGRISILDD-----DIVDNEMAAILOGFRS 382
DB 2184 LFYLSAKAFIDFLAIEKR---KGKSEFLDWSSGGRFYEDTLTIDS-----YWRRI 2233
QY 383 MIEQFNVNPAATAKELQAMENQUTLANSDDLVGADG---ELPAETQAIKDALQALQKPS 438
DB 2234 VASRGRNGRTISYBALDGPRASTIVPTSHSTSPGYTILDVANAMLFVGLGTGLKRA 2293
QY 439 ADGLAT---AMGOVAFAAAKVG-----GGSAG-----TAGYQNM----- 470
DB 2294 AVRYITFTGCGEYTFPNKRPDILGMNFRKEDGDKGCTVSPQVEDSBGTIOFGEGEYALVS 2353
QY 471 -----VQKYLTAPSTSTSSSYAA-----LSDGYSAVKTLSLNSPSR 509
DB 2354 RPIRMYPNISTVMEKFTF-FSSSLMLYLTATRLDRFPMSVELTDGHT-----KSYDUG 2406
QY 510 SGVSOAIS-QTANPALRSYSRSRGISOGKSA--DASORAAETIVRDS 554
DB 2407 SGMASVSYSONHNDGKKKSFTELRIOQANISIVDITDQNEINATSS 2454

RESULT 35
US-08-125-077-4
; Sequence 4, Application US/08125077
; Patent No. 5872231
; Patent No. 5872231 5840863
; GENERAL INFORMATION:
; APPLICANT: Engvall, Eva
; APPLICANT: Leivo, Ilmo
; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
; TITLE OF INVENTION: Fragments and Uses Thereof
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

1      RESULT 35
2      US-08-125-077-4
3      ; Sequence 4, Application US/08125077
4      ; Patent No. 5872231
5      ; Patent No. 5872231 5840863
6      ; GENERAL INFORMATION:
7      ; APPLICANT: Engvall, Eva
8      ; APPLICANT: leiwo, Ilmo
9      ; TITLE OF INVENTION: Nucleic Acids Encoding Merosin, Merosin
10     ; TITLE OF INVENTION: Fragments and Uses Thereof
11     ; NUMBER OF SEQUENCES: 23
12     ; CORRESPONDENCE ADDRESS:
13     ; ADDRESSEE: Campbell and Flores
14     ; STREET: 4370 La Jolla Village Drive, Suite 700
15     ; CITY: San Diego
16     ; STATE: California
17     ; COUNTRY: USA
18     ; ZIP: 92122
19     ; COMPUTER READABLE FORM:
20     ; MEDIUM TYPE: Floppy disk
21     ; COMPUTER: IBM PC compatible
22     ; OPERATING SYSTEM: PC-DOS/MS-DOS
23     ; SOFTWARE: Patentln Version #1.0, Version #1.25
24     ; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/125,077
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US PCT/US 94/10730
; FILING DATE: 21-SEP-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/472,319
; FILING DATE: 30-JAN-1990
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/919,951
; FILING DATE: 27-JUL-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LA 9721
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3111 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-125-077-4

```

Query Match 5.3%; Score 170.5; DB 2; Length 3111;

Best Local Similarity 19.9%; Pred. No. 0.0069; Mismatches 259; Indels 157; Gaps 26;

```

; 13 ESYSSNOSNNPIINOIASNSETKSTKASEASPSASSSSVSWFLSSAKMALISLRDA 72
; 1858 EDIOTKLPSEELNIDKIDLSOEIKDKRLAEKVSQASHA-----A 1899
; 73 ILKNNSSPTSLSOLEASTSTVTYRAADYDEAKSNFTAKSGLENAKTLAEYETKMA 132
; 1900 QLDSSAVLDGI--LDEAKNISFNATPAFAAYSNIKDYIDEAKVNAEADLHAETKLA 1957
; 133 -----DLMAALQ-----DMERLANSDESNHTEEVNNIKKALE----- 165
; 1958 TGRGLIKEAKGCLQKSPFLMEAKKLANDVENE--DHLNGLKTIENADARNGDLR 2015
; 166 AQRDTIDKLN-----KLVTLQONKSLTEVLTQTTSDAQIPAINSOLEINKNSADQI 217
; 2016 TLNDTGLKISAIENDPAKLAQAVKAKARQANDTAK--DVLAQITELHQNDLGLKKNYKL 2073
; 218 IKLERINIEYEAULTNAGE-VIKASEAGIK-LGQALQSIYVADQSQAAVLAQOONNS 275
; 2074 ADSVAKTN--AVVKDPSKNKTIADADATVKNLEQADRLID---KLKPIKELEDNL 2124
; 276 PDNIAATKELIDAETKVNELKQEHGTLDSPLVKKAEQISOAQKDIOEIKPSGSDIPI 335
; 2125 KKNISEKELINARKQANIKVSSSGDCITRYKPEIKKGSNNIYVAVKTRAVAD-NL 2183
; 336 VGPSGAASAGSAGALKSSNSNGRISLLD-----DVDNEMAATALOGFRS 382
; 2184 LFLYLSAKFIDFLAIEMR---KGVSFLMDVSGVGRVEYPLDTIDDS-----WYRI 2233
; 383 MIOGFANNATAKELQAMEAQLTAMSDQLVADG---ELPAEIOAKIKALQALQOPS 438
; 2234 VASRTGNGTISYRALDGPAPLSTPHSTSPPGYTIIDVDANAMLVFGGLGKLKKAD 2293
; 439 ADGLAT---AMGOVAFAAKVG-----GSGAG-----TACTYOMN----- 470
; 2294 AVRYITFTGCMGETYFPNKRKIGLMNFRERKGDCKGCTVSPQVESEBGTIQPDGQYALVS 2353
; 471 -----VKQYKTAESSSTSSSYAA-----LSDQSYAYKTLNLSYSESR 509
; 2354 RPIRWYNISTVMEKFKRT-FSSSALLMYLATRDLRDFMSVELTGH-----KVSVDLG 2406
; 510 SGVQSAIS-OTANPALSRVSRSGIESQGRSA--DASQRAEFTIVRS 554
; 2407 SGHASYVSNQNHNDGKWKSFLLSRIOKQANISYVIDIDTNOENIATSS 2454

```

```

; RESULT 36
; 5352450-2
; Patent No. 5352450
; APPLICANT: KOGA, TOSHIHIKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;
; SHIBUYA, KOJI; OHTA, HIROAKA
; TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
; CARIES AND VACCINAL COMPOSITIONS FOR DENTAL CARIES USED AS NASAL
; DROP
; NUMBER OF SEQUENCES: 2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/529,602
; FILING DATE: 29-MAY-1990
; SEQ ID NO: 2
; LENGTH: 1565
; 5352450-2

```

Query Match 5.3%; Score 169.5; DB 6; Length 1565;

Best Local Similarity 18.6%; Pred. No. 0.0031; Mismatches 267; Indels 117; Gaps 21;

```

; 10 SCPEYSSNOSNNPIINOIASNSETKSTKASEASPSASSSSVSWFLSSAKMALISL 69
; 32 AGQKVFADETTTTSDVDKRVGTQGNPATQLPEAGSAS-----KEAQSQTKL 82
; 70 RDAIILKNNSSPTSLSOLEASTSTVTYRAADYDEAKSNFTAKSGLENAKTLAEYET 129
; 83 ERQAVHTIEVKTDLDOAAKDAKSAGYVNV--QDADYKCTVTPPEAVQKETEIEKEDT 140
; 130 KMAIDLMAALQDMERLANSDESN--NHTEEVNNIKKALEAKOKTIDK-----LNKLVY 179
; 141 KQA-----EDIKTTIDQTKSDVAHAHEAEVAKIKAKNOATKEQYEDMAHKAKEVERINA 194
; 180 LQONKNSLTFVLKTTSDAQIPAINSOLEINKNSAQOIIKLERONISYEAVLTNAGEVY 239
; 195 ANASTATPAEAKLAQYQAD-LAAVQ-----KTNAANO-----AAIQKALAAVQAE 239
; 240 KASEGKIKLQALQSIYVADQSQAAVLAQOONNSPDNIAATKELIDAETKVNELKOE 299
; 240 KRYQENAAKAAVDPYVAVANNNAKNTIEAANBEIKRNTAKAEV---ETKLAQYQAE 295
; 300 HTGLTDSPLVKKAEQ--ISQAKDIOEIKPSGSDIPIVGPSSAASAGSAGALKSSNN 357
; 296 LKRYQENAAANEADYQAKLTAYOTELARYOKANAD-----AKATYEAAVAANNAKN 346
; 358 SGRIISLLDQVDMEMALALQGRSMIEGFNVNPPATK-ELQAMEAQLTAMSDQLVAD 416
; 347 A-----ALTAENTFAIKERNENAKATYEALQYEDLAAVKKANANNE 389
; 417 GELPAEIOAIKDALQALQOPASADGLATAMGOVAFAAKVGSGSAGTAGTVQ----- 468
; 390 ADYQAKLTAYQTELARQK-----ANMDAKAYEAAVAANNAALTAENTATKKRN 442
; 469 MNVQOLY--KTAESSSTSSSYAALSDGYSAVYTLNLSYSESRSGVQSAISQGANPALSR 526
; 443 ADAKADYEAKLARYQADLARYQKDLAD---YVKKLAKYEDBOTSIIKALAE-----LEK 493
; 527 SVSRSGIESGGRADDSQRAEFTIVRD-----SQTIGVYRLQVILSMTIYSNPQ 579
; 494 HKMEDG-----NLTEPSAQNLYVDLEPNANLSLTTDGKFLKASAVDDAFSKSTSKAK 545
; 580 ANOETIMQ 587
; 546 YDQ-KILQ 552

```

RESULT 37
US-09-541-782-6
Sequence 6, Application US/09541782
Patent No. 6284480
GENERAL INFORMATION:
APPLICANT: Nislow, Corey


```

Oy 327 -----KPSGSD-----IPYGP5-----GSAASAGSAGALKSSNNSGRISLL 365
Db 960 GATVSEDDGKXDGTGIVTAKTVIDAVNKSGWRVTGEGATAETGATAVNAGNAETVTSCTS 1019
Oy 366 DDVNDEMAIAL-----QCFRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGELPA 421
Db 1020 VNFKNNGNATTATVSKDNGNINVKYDVNVGDGLKIGDDKKIVADTTLT--VTGKVSYP 1077
Oy 422 EIQAIKDALQALQPSADGLATAMQOYAFAAKAVGGSGAGTACTYQNNVKQLYKTAESS 481
Db 1078 GANSVNN-----NKKLVNAEGLTATALNLSMTAKADKYADGESEGETDOEVKAGDKVTFKA 1133
Oy 482 TSSSSYAAALSD-GYSAYKTLNLSYSESRSQVQSAISOT-----ANPALSR- 526
Db 1134 GKNLKVKQSEKDFYSLQDTLTGLTSTLGTANGRNDGTIVINKDGLTITLANGAAGT 1193
Oy 527 -----SVRSQIESQSRASQRAAETIYRDSQTLGDVYSRLQVLDLSIMSTIVSNP 578
Db 1194 DASNGNTISVFKDGI-SAGNKEITNVKSALKTKYKDTQNTAD-----ETQDKREFHAAYKN- 1246
Oy 579 QANQEEIMOKLTASIS 594
Db 1247 -ANEVEFYGKNGATVS 1261

```

Search completed: March 12, 2003, 12:23:06
 Job time : 77 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 09:40:57 ; Search time 66 Seconds
(without alignments)
1332.506 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHMHESGPESVSSNQS.....PAFIQYLVNIAISLFGYLS 660

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3204	100.0	660	22 AAU38921	C. trachomatis CT8
2	2943	91.9	647	23 ABB94283	Chlamydia trachoma
3	1690	52.7	361	21 AAB13695	Chlamydia sp. prot
4	1690	52.7	361	22 AAG83263	Protein encoded by
5	1690	52.7	361	23 ABB94234	Chlamydia trachoma
6	1517	47.3	331	20 AAY37572	Chlamydia trachoma
7	1435	44.8	350	20 AAY37571	Chlamydia trachoma
8	894	27.9	651	22 AAY71954	Chlamydia pneumonia
9	894	27.9	651	22 AAU38899	C. pneumoniae CT62
10	894	27.9	651	23 ABB90532	Chlamydia pneumonia

11	894	27.9	651	23 ABB94277	Chlamydia pneumonia
12	830.5	25.9	583	21 AAY71955	5'-truncated Chlam
13	453	14.1	452	21 AAY71956	3'-truncated Chlam
14	453	14.1	490	21 AAY71957	C. pneumoniae 76 k
15	451	14.1	478	20 AAY35358	Chlamydia pneumonia
16	445.5	13.9	212	20 AAY35357	Chlamydia pneumonia
17	256	8.0	2344	22 AAU37120	Staphylococcus aur
18	254	7.9	6281	22 AAU37403	Staphylococcus aur
19	251	7.8	1463	23 AAE20110	Staphylococcus aur
20	249.5	7.8	5024	22 AAG82935	Lactobacillus rham
21	248.5	7.8	5024	22 AAG82935	S. epidermidis ope
22	247	7.7	3158	22 AAU37018	Staphylococcus epi
23	244.5	7.6	2659	23 AAU75489	Staphylococcus aur
24	239.5	7.5	2434	22 AAU34339	S. aureus antigen
25	237	7.4	2137	23 ABB39618	Staphylococcus epi
26	236	7.4	1215	22 AAU34412	Staphylococcus epi
27	236	7.4	1269	22 AAU34338	Staphylococcus aur
28	235.5	7.4	2437	22 AAU34338	Staphylococcus aur
29	235	7.3	2285	20 AAM98149	Bacillus subtilis
30	234.5	7.3	1029	22 AAU34389	Staphylococcus aur
31	234.5	7.3	1048	22 AAU37490	Staphylococcus aur
32	234	7.3	560	22 AAU37464	Staphylococcus aur
33	234	7.3	1063	23 ABB54168	Lactococcus lactis
34	233	7.3	2025	22 AAU34207	Staphylococcus aur
35	228	7.1	2086	22 AAU34143	Staphylococcus aur
36	228	7.1	5795	22 AAU37017	Staphylococcus aur
37	227.5	7.1	1095	22 AAG83030	S. epidermidis ope
38	222	6.9	2478	22 AAU34320	Staphylococcus aur
39	222	6.9	2478	22 AAU34320	Staphylococcus aur
40	219.5	6.9	1833	22 ABB71141	Drosophila melanog
41	219	6.8	3696	23 ABB40235	Staphylococcus epi
42	217	6.8	2368	22 AAU34139	Staphylococcus aur
43	217	6.8	2368	22 AAU36796	Staphylococcus aur
44	214	6.7	1072	23 ABB54963	Staphylococcus aur
45	213	6.6	2186	22 AAU37320	Lactococcus lactis

ALIGNMENTS

RESULT 1
AAU38921
ID AAU38921 standard; Protein: 660 AA.
AC AAU38921;
XX 16-JAN-2002 (first entry)
DT
XX
DE C. trachomatis CT875 protein.
XX
KW Chlamydia; sexually transmitted disease; PID; antibacterial;
KW pelvic inflammatory disease; antigen; trachoma; gynecological;
KW acute respiratory tract infection; atherosclerosis; male infertility;
KW coronary heart disease.
XX
OS Chlamydia trachomatis serovar E.
XX
PN WO200181379-A2.
PD 01-NOV-2001.
XX
PF 23-APR-2001; 2001WO-US13081.
XX
PR 21-APR-2000; 2000US-198853P.
XX 20-JUL-2000; 2000US-219752P.
PA (CORI-) CORIXA CORP.
XX Bhatia A, Probst P, Stromberg EJ;
XX WPI; 2001-616771/71.
DR N-PSDB; AAS56995.
XX

PT New polynucleotide for treating Chlamydia infections encodes a
PT polynucleotides containing an immunogenic portion of a Chlamydia
PT antigen -
XX
XX
PS Claim 3: Page 205-207; 208pp; English.
XX
CC The invention relates to isolated polynucleotide encoding at least
CC a partial Chlamydia protein which is an antigenic fragment, or the
CC complements, fragments, homologues and variants, and antibodies
CC raised against the antigenic proteins (or fragments). The nucleic
CC acids, proteins and antibodies are used to diagnose and treat Chlamydia
CC infections (e.g. a sexually transmitted disease, pelvic inflammatory
CC disease (PID), acute respiratory tract infection, trachoma,
CC atherosclerosis and coronary heart disease) in a patient, and in
CC the treatment of male infertility. The compounds of the invention are
CC also useful for detecting the presence of Chlamydia in a patient, and
CC stimulating and/or expanding T cells specific for a Chlamydia protein.
CC The present sequence represents a Chlamydia antigen.
XX
SQ Sequence 660 AA:

Query Match 100.0%; Score 3204; DB 22; Length 660;
Best Local Similarity 100.0%; Pred. No. 4.3e-185;
Matches 660; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHMHMHMEGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSL 60
DB 1 MHMHMHMEGPESSVSSNMPININGQIASNSETKESTKASEASPSASSVSMSFSL 60
QY 61 SAKNALISLRDALINKNSPTDLSQLEASTSTYTRVAAKYDEAKSNFTAKSGLEN 120
DB 61 SAKNALISLRDALINKNSPTDLSQLEASTSTYTRVAAKYDEAKSNFTAKSGLEN 120
QY 121 AKTAEYERKMDLMAALDMERLANSPPSNHTEEVNNTKKALEOKDTIDKLNKLVTL 180
DB 121 AKTAEYERKMDLMAALDMERLANSPPSNHTEEVNNTKKALEOKDTIDKLNKLVTL 180
QY 181 QONKSLTEVLKTTDSADQIPAINSOLEINKNSADQIINDLERONTSYEAVLTNAGEVIR 240
DB 181 QONKSLTEVLKTTDSADQIPAINSOLEINKNSADQIINDLERONTSYEAVLTNAGEVIR 240
QY 241 ASSEAIKIKGQALQSTIVDAGDOSQAAYVLAQOONNSPDNIAATKELIDAEETKYNELKOEI 300
DB 241 ASSEAIKIKGQALQSTIVDAGDOSQAAYVLAQOONNSPDNIAATKELIDAEETKYNELKOEI 300
QY 301 TGLTDSPLVKKAEQISOAKDIOEIKPSSGSDPIVGPSSAASAGALSSNNSGR 360
DB 301 TGLTDSPLVKKAEQISOAKDIOEIKPSSGSDPIVGPSSAASAGALSSNNSGR 360
QY 361 ISLLDDVDNEMAIALOGFRSMIEQFNVPNNPATAKELQAMPAQLTAMSDQVAGDELP 420
DB 361 ISLLDDVDNEMAIALOGFRSMIEQFNVPNNPATAKELQAMPAQLTAMSDQVAGDELP 420
QY 421 AETQATKDALQALKOPSSADGLATAMGOVAFPAAKYGGSGAGTAGYVONNVKQLYTAFS 480
DB 421 AETQATKDALQALKOPSSADGLATAMGOVAFPAAKYGGSGAGTAGYVONNVKQLYTAFS 480
QY 481 STSSSSYAALSDGYSAYTTLNLSYESRSGVQSAISQTPANPLASVSFSGIESQGRSA 540
DB 481 STSSSSYAALSDGYSAYTTLNLSYESRSGVQSAISQTPANPLASVSFSGIESQGRSA 540
QY 541 DASQRAETIVRDSQTLGVDYSRLQVLSLMTSTIVSNPQANDEIMOKLTASISKAPOFG 600
DB 541 DASQRAETIVRDSQTLGVDYSRLQVLSLMTSTIVSNPQANDEIMOKLTASISKAPOFG 600
QY 601 YPAVONASDLSQFAAQOLEREFYDGERSLAESQENAFKROPAPITQOVLVNIASLFSGYLS 660
DB 601 YPAVONASDLSQFAAQOLEREFYDGERSLAESQENAFKROPAPITQOVLVNIASLFSGYLS 660

RESULT 2
ABB94283
ID ABB94283 standard; Protein; 647 AA.

XX ABB94283;
AC 05-JUN-2002 (first entry)
XX
DT
XX
DE Chlamydia trachomatis protein sequence SEQ ID NO:436.
XX
XX Chlamydia infection; Chlamydia; vaccine; detection; diagnosis;
KW antigen; antibacterial; immunostimulant; immune response;
KW Chlamydia-specific T-cell response.
XX
OS Chlamydia trachomatis.
XX
PN WO200208267-A2.
XX
PF 20-JUL-2001; 2001WO-US23121.
XX
PR 20-JUL-2000; 2000US-0620412.
PR 23-APR-2001; 2001US-0841132.
XX
PA (CORI-) CORIXA CORP.
XX
PI Fling SP, Skeiky YAM, Probst P, Bhatia A;
XX WPI; 2002-179901/23.
XX
PT Novel compositions comprising Chlamydia Cap1 protein and its use in the
PT treatment of Chlamydia infection -
XX
PS Claim 34; Page 391-393; 537pp; English.

CC The present invention describes compositions comprising a Chlamydia Cap1
CC protein and methods for the diagnosis and therapy of Chlamydia infection.
CC Chlamydia DNA and protein sequences from the present invention can have
CC antibacterial and immunostimulant activities, and can be used in
CC vaccines. Compounds from the present invention can be used for eliciting
CC an immune response, specifically stimulating a Chlamydia-specific T-cell
CC response or inhibiting the development of a Chlamydia infection in an
CC animal. Methods from the present invention can be used for detecting the
CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
CC specific for a Chlamydia protein; and for treatment of a Chlamydia
CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 647 AA:

Query Match 91.9%; Score 2943; DB 23; Length 647;
Best Local Similarity 93.9%; Pred. No. 2.3e-169;
Matches 613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;

QY 8 MESGPESSVSSNOSNMPIINGQIASNSETKESTKASEASPSASSVSMSFSLSSAKNALI 67
DB 1 MESGPESSVSSNOSNMPIINGQIASNSETKESTKASEASPSASSVSMSFSLSSAKNALI 60
QY 68 SLRDAIINKNSPTDLSQLEASTSTYTRVAAKYDEAKSNFTAKSGLENAKTAEY 127
DB 61 SLRDAIINKNSPTDLSQLEASTSTYTRVAAKYDEAKSNFTAKSGLENAKTAEY 120
QY 128 ETKMADLMAALQMERLANSPPSNHTEEVNNTKKALEOKDTIDKLNKLVTLQONKSL 187
DB 121 ETKMADLMAALQMERLANSPPSNHTEEVNNTKKALEOKDTIDKLNKLVTLQONKSL 174
QY 188 TEVLKTTDSADQIPAINSOLEINKNSADQIINDLERONTSYEAVLTNAGEVIRKASEAGI 247
DB 175 KETLTTTDSADQIPAINSOLEINKNSADQIINDLERONTSYEAVLTNAGEVIRKASEAGI 234
QY 248 KLGQALQSTIVDAGDOSQAAYVLAQOONNSPDNIAATKELIDAEETKYNELKOEHTTGLTDSF 307
DB 235 KLGQALQSTIVDAGDOSQAAYVLAQOONNSPDNIAATKELIDAEETKYNELKOEHTTGLTDSF 294
QY 308 LVKKAEEQISOAKDIOEIKPSSGSDPIVGPSSAASAGALSSNNSGRISLLDD 367

```

|||||
Db 295 LVKKAEQISQAKDIOEIKPSSGSDIPYVPGSSAASAGALKSSNNSGRISLLDD 354
OY 368 VDNEMAAIALOGFRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGELPAEIQAK 427
Db 365 VDNEMAAIAMOGFRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGELPAEIQAK 414
OY 428 DALAALQKPSADGLATAMGQVAFPAAKVGGSGAGTACTAGTQVMNVKQLYTAFTSSSSY 487
Db 415 DALAALQKPSADGLATAMGQVAFPAAKVGGSGAGTACTAGTQVMNVKQLYTAFTSSSSY 474
OY 468 AAALSDGYSAKYKTLNLSYSESRGVQSAISQTPANPALSRVSRSRGIESQGRSADASQRAA 547
Db 475 AAALSDGYSAKYKTLNLSYSESRGVQSAISQTPANPALSRVSRSRGIESQGRSADASQRAA 534
OY 548 ETIVDSQTLGIVSRLOYLDLSMTSTIVSNPQANQEIIMOKLTASISKAPQGYPAVONS 607
Db 535 ETIVDSQTLGIVSRLOYLDLSMTSTIVSNPQANQEIIMOKLTASISKAPQGYPAVONS 594
OY 608 ADSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFTQOVLVNIASLFSGYLS 660
Db 555 ADSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFTQOVLVNIASLFSGYLS 647

RESULT 3
AAB13695
ID AAB13695 standard; Protein; 361 AA.
XX
AC AAB13695;
XX
DT 02-FEB-2001 (first entry)
XX
DE Chlamydia sp. protein # 6.
XX
KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial.
XX
OS Chlamydia sp..
XX
PN WO200034483-A2.
XX
PD 15-JUN-2000.
XX
PE 08-DEC-1999; 99WO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
XX (CORI-) CORIXA CORP.
XX
PI Probst P, Bhatia A, Skelky YAW, Fling SP, Jen S, Stromberg EJ;
XX
XX WPI; 2000-431303/37.
XX
DR 15-JUN-2000.
XX
PF Isolated polypeptide for diagnosis and treatment of Chlamydia infection
XX PT comprises immunogenic portion of Chlamydia antigen, which comprises
XX PT amino acid sequence encoded by polynucleotide sequence -
XX
PS Disclosure: Pages 251-253; 256pp; English.
XX
XX The present invention relates to new nucleic acid sequences and the
XX CC proteins encoded by the nucleic acid sequences. The encoded proteins
XX CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
XX CC proteins are useful for the serodiagnosis and treatment of Chlamydia
XX CC infection. Chlamydiae are intracellular bacterial pathogens that are
XX CC responsible for a wide variety of human infections. C. trachomatis
XX CC infection is one of the most common sexually transmitted diseases and can
XX CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
XX CC and infertility. Trachoma due to ocular infection with C. trachomatis is

```

```

CC the leading cause of preventable blindness worldwide. C. pneumonia is a
CC major cause of acute respiratory tract infections in humans and is also
CC thought to play a role in the pathogenesis of atherosclerosis and
CC coronary heart disease. The present sequence is a protein isolated in the
CC present invention.
XX
SQ Sequence 361 AA;
XX
Query Match 52.7%; Score 1690; DB 21; Length 361;
Best Local Similarity 97.0%; Pred. No. 3,9e-94;
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
OY 300 HTGLDPSLVKKAEQISQAKDIOEIKPSSGSDIPYVPGSSAASAGALKSSNNSG 359
Db 1 HOEIDSPLYKKAEQINQAOODIOTITPSGLDIPYVPGSSAASAGALKSSNNSG 60
OY 360 RISLLDDVDNEMAAIALOGFRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGEL 419
Db 61 RISLLDDVDNEMAAIAMOGFRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGEL 120
OY 420 PAEIQAIKDALQALQKPSADGLATAMGQVAFPAAKVGGSGAGTACTAGTQVMNVKQLYTAF 479
Db 121 PAEIQAIKDALQALQKPSADGLATAMGQVAFPAAKVGGSGAGTACTAGTQVMNVKQLYTAF 180
OY 480 SSTSSSYAAALSDGYSAKYKTLNLSYSESRGVQSAISQTPANPALSRVSRSRGIESQGRS 539
Db 181 SSTSSSYAAALSDGYSAKYKTLNLSYSESRGVQSAISQTPANPALSRVSRSRGIESQGRS 240
OY 540 ADASQRAETIVRDSQTLGDVYSRLQYLDLSMTSTIVSNPQANQEIIMOKLTASISKAPQF 599
Db 241 ADASQRAETIVRDSQTLGDVYSRLQYLDLSMTSTIVSNPQANQEIIMOKLTASISKAPQF 300
OY 600 GYPVONSDSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFTQOVLVNIASLFSGYL 659
Db 301 GYPVONSDSLQKFAQLEREFVDSERSLAESQENAFRRKQPAFTQOVLVNIASLFSGYL 360
OY 660 S 660
Db 361 S 361

RESULT 4
AAG83263
ID AAG83263 standard; Protein; 361 AA.
XX
AC AAG83263;
XX
DT 05-SEP-2001 (first entry)
XX
DE Protein encoded by Chlamydia trachomatis clone CT622.
XX
KW Chlamydia; vaccine; infection; fusion protein; antigen;
KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW acute respiratory tract infection; Cap1; CF529; OMCB;
KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS Chlamydia trachomatis.
XX
PN WO200140474-A2.
XX
PD 07-JUN-2001.
XX
PE 04-DEC-2000; 2000WO-US32919.
XX
PR 03-DEC-1999; 99US-0454684.
PR 19-APR-2000; 2000US-0556877.
PR 20-JUN-2000; 2000US-0598419.
XX
XX (CORI-) CORIXA CORP.
XX
PI Probst P, Bhatia A, Skelky YAW, Fling SP, Scholler J;
XX
XX WPI; 2001-374831/39.

```

XX Chlamydia polypeptides and fusion proteins useful for preventing pelvic
PT inflammatory disease, trachoma, acute respiratory tract infections,
PT atherosclerosis and heart disease -
XX
PS Claim 66; Page 257-259; 295pp; English.
XX
CC The present sequence is provided in a specification relating to
CC compounds and methods for the treatment and diagnosis of chlamydial
CC infection. The compounds provided include polypeptides and fusion
CC proteins comprising immunogenic portions of Chlamydia antigens and
CC and DNA sequences encoding such polypeptides. They are useful for
CC vaccinating against chlamydial infection, which causes pelvic
CC inflammatory disease, trachoma, acute respiratory tract infections,
CC atherosclerosis and heart disease.
XX
SQ Sequence 361 AA:

Query Match 52.7%; Score 1690; DB 22; Length 361;
Best Local Similarity 97.0%; Pred. No. 3.9e-94;
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 300 HTGLTSPLYKKAEEQISQAKDIOETKPSGSDPIVGPBSAASAGALKSSNNG 359
| : |||||||||:|||||: || |||||||||:|||||||
Db 1 HOEIA DSPLYKKAEEQINQAOQDIOITTPSGLDIPVGPBSAASAGALKSSNNG 60

QY 360 RISLLDDVDNEMAAIATLQFRSMIEQFNPNPATKKELOAMEAOULTAMSDQLVGADGEL 419
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 61 RISLLDDVDNEMAAIATLQFRSMIEQFNPNPATKKELOAMEAOULTAMSDQLVGADGEL 120

QY 420 PAETIOAIKDALQALQKPSADGLATAMGOVAFAPAAKVGSGSAGTACTVQNNVQLYKTA 479
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 121 PAETIOAIKDALQALQKPSADGLATAMGOVAFAPAAKVGSGSAGTACTVQNNVQLYKTA 180

QY 480 SSTSSSSYAAALSDGYSAVYTLNLSYSESRGVSQSAISQANPALSRVSRSIEQGRS 539
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 181 SSTSSSSYAAALSDGYSAVYTLNLSYSESRGVSQSAISQANPALSRVSRSIEQGRS 240

QY 540 ADASORAETIVRDQSLGIVSRLOVLSLMTIYSNPQANOEIMOKLTASISKAPOF 599
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 241 ADASORAETIVRDQSLGIVSRLOVLSLMTIYSNPQANOEIMOKLTASISKAPOF 300

QY 600 GYPAVONSADSLQKFAAQLEREFVDGERSLAESQENAFKOPAFIOQVLVNIASLFSGYL 659
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 301 GYPAVONSADSLQKFAAQLEREFVDGERSLAESQENAFKOPAFIOQVLVNIASLFSGYL 360

QY 660 S 660
|
Db 361 S 361

RESULT 5
ABB94234
ID ABB94234 standard; Protein; 361 AA.

AC ABB94234;
DT 05-JUN-2002 (first entry)

DE Chlamydia protein sequence SEQ ID NO:299.

XX Chlamydia infection; Chlamydia; vaccine; detection; diagnosis;
KW antigen; antibacterial; immunostimulant; immune response;
XX Chlamydia-specific T-cell response.

OS Chlamydia sp.

PN WO200208267-A2.

PD 31-JAN-2002.

PF 20-JUL-2001; 2001WO-US23121.

XX

PR 20-JUL-2000; 2000US-0620412.
PR 23-APR-2001; 2001US-0841132.
XX
PA (CORI-) CORIAX CORP.
PI Filing SP, Skeiky YAW, Probst P, Bhatia A;
XX
DR WPI; 2002-179901/23.
XX
PT Novel compositions comprising Chlamydia Cap1 protein and its use in the
PT treatment of Chlamydia infection -
XX
PS Disclosure; Page 284-285; 537pp; English.
XX
CC The present invention describes compositions comprising a Chlamydia Cap1
CC protein and methods for the diagnosis and therapy of Chlamydia infection.
CC Chlamydia DNA and protein sequences from the present invention can have
CC antibacterial and immunostimulant activities, and can be used in
CC vaccines. Compounds from the present invention can be used for eliciting
CC an immune response, specifically stimulating a Chlamydia-specific T-cell
CC response or inhibiting the development of a Chlamydia infection in an
CC animal. Methods from the present invention can be used: for detecting the
CC presence of Chlamydia in a patient; to stimulate and/or expand T cells
CC specific for a Chlamydia protein; and for treatment of a Chlamydia
CC infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent
CC sequences used in the exemplification of the present invention.
XX
SQ Sequence 361 AA:

Query Match 52.7%; Score 1690; DB 23; Length 361;
Best Local Similarity 97.0%; Pred. No. 3.9e-94;
Matches 350; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 300 HTGLTSPLYKKAEEQISQAKDIOETKPSGSDPIVGPBSAASAGALKSSNNG 359
| : |||||||||:|||||: || |||||||||:|||||||
Db 1 HOEIA DSPLYKKAEEQINQAOQDIOITTPSGLDIPVGPBSAASAGALKSSNNG 60

QY 360 RISLLDDVDNEMAAIATLQFRSMIEQFNPNPATKKELOAMEAOULTAMSDQLVGADGEL 419
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 61 RISLLDDVDNEMAAIATLQFRSMIEQFNPNPATKKELOAMEAOULTAMSDQLVGADGEL 120

QY 420 PAETIOAIKDALQALQKPSADGLATAMGOVAFAPAAKVGSGSAGTACTVQNNVQLYKTA 479
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 121 PAETIOAIKDALQALQKPSADGLATAMGOVAFAPAAKVGSGSAGTACTVQNNVQLYKTA 180

QY 480 SSTSSSSYAAALSDGYSAVYTLNLSYSESRGVSQSAISQANPALSRVSRSIEQGRS 539
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 181 SSTSSSSYAAALSDGYSAVYTLNLSYSESRGVSQSAISQANPALSRVSRSIEQGRS 240

QY 540 ADASORAETIVRDQSLGIVSRLOVLSLMTIYSNPQANOEIMOKLTASISKAPOF 599
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 241 ADASORAETIVRDQSLGIVSRLOVLSLMTIYSNPQANOEIMOKLTASISKAPOF 300

QY 600 GYPAVONSADSLQKFAAQLEREFVDGERSLAESQENAFKOPAFIOQVLVNIASLFSGYL 659
|||||||:|||||||:|||||||:|||||||:|||||||:|||||||
Db 301 GYPAVONSADSLQKFAAQLEREFVDGERSLAESQENAFKOPAFIOQVLVNIASLFSGYL 360

QY 660 S 660
|
Db 361 S 361

RESULT 6
AA37572
ID AA37572 standard; Protein; 331 AA.

AC AA37572;
DT 07-OCT-1999 (first entry)

DE Chlamydia trachomatis cellular envelope protein.
KW vaccine; eye disease; conventional trachoma; nonendemic trachoma;

XX

KM paratrachoma: inclusion conjunctivitis; genital disease; perithenitis;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KM Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX Chlamydia trachomatis.
OS
XX
XX MO9928475-A2.
PN
XX
XX 10-JUN-1999.
PD
XX
XX 27-NOV-1998; 98MO-IB01939.
PF
XX
XX 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
XX (GEST) GENSET.
PA
XX
XX Griffiths R;
PI
XX
XX WPI: 1999-371125/31.
DR
XX
XX Genome sequence of Chlamydia trachomatis
PT
XX
XX Disclosure; Page 1227-1228; 1755pp; English.
PS
XX
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nongonococcal urethritis,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perithenitis, Bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
XX
SQ Sequence 331 AA:
Query Match 47.3%; Score 1517; DB 20; Length 331;
Best Local Similarity 99.4%; Pred. No. 9.4e-84;
Matches 315; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

AC AAY37571;
XX
XX 07-OCT-1999 (first entry)
DT
XX
XX Chlamydia trachomatis surface exposed protein.
DE
XX
XX
XX Vaccine; eye disease; conventional trachoma; nongonococcal urethritis;
KM paratrachoma: inclusion conjunctivitis; genital disease; perithenitis;
KM nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KM Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
XX Chlamydia trachomatis.
OS
XX
XX MO9928475-A2.
PN
XX
XX 10-JUN-1999.
PD
XX
XX 27-NOV-1998; 98MO-IB01939.
PF
XX
XX 04-NOV-1998; 98US-0107077.
PR 28-NOV-1997; 97FR-0015041.
PR 17-DEC-1997; 97FR-0016034.
XX
XX (GEST) GENSET.
PA
XX
XX Griffiths R;
PI
XX
XX WPI: 1999-371125/31.
DR
XX
XX Genome sequence of Chlamydia trachomatis
PT
XX
XX Disclosure; Page 1226-1227; 1755pp; English.
PS
XX
XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye
CC diseases such as conventional trachoma, nongonococcal urethritis,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perithenitis, Bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.
XX
XX
SQ Sequence 350 AA:
Query Match 44.8%; Score 1435; DB 20; Length 350;
Best Local Similarity 89.5%; Pred. No. 8.9e-79;
Matches 299; Conservative 17; Mismatches 16; Indels 2; Gaps 1;

Dd	125	TSLDIAQIALVSLQDAVINIKDTATD-----EET-----AIAAEWETKKNADAVKVGQ	173
Qy	174	LNKIVTTLQNNKSKSLTEVLEKTTDSADQIPAINLSQLEINRNSADQIIKDLERQNI---SYEA	230
Dd	174	ITELAKVASYDQAIJLDSLGKLTSPDLLQALLQISVAANNKKAELLKEMQDNVVPGRKTPA	233
Qy	231	VLTNAGEVTKASSESGIGILGQALOSTIVADGDSQAAYVLOAQONNSPDNIATKELIDAAE	290
Dd	234	IAQSLVDPDTATATDIEKDGNAIRPAYFAGQNASGAVENAKSNNSISNDSKAKAIATATK	293
Qy	291	TKVNELOKEHGLTSPSPVKKKEQISOAQKDIOEIKPS-GSDIPTGPS-GSAASAGSA	348
Dd	294	TOIAEAQKK---FPDSPTILOEAEQVYTOAEKDKNIKFPADGSDVNPGETTVGGSKQGGSS	350
Qy	349	AGALKSSNNSGRISILLDDVDNEMAAIALQGFRRMIEQFNVNNPATAKELQAMEKQLTAM	408
Dd	351	IGSI-----RVSMLLDDAEKETASILMSGFRQMIHMFENENPDSQAQOJELAAQARA	403
Qy	409	SDQIVGADGELPAETQAIKDALQAL-KQPSADGLATAMGOVAFPAARVGGSGAGTAGTV	467
Dd	404	K---AAGDSSAAALADQKALEALGKAGQOQGILNGLQIASAAVVSAGVPPAAASSI	460
Qy	468	QMNVMOLKKTAFSSSSSSSYAAALSDGSAVYKTLMSLSESS-GVQASISGTANPALSR	526
Dd	461	GSSVQQLKKT--SKTSGDYTKQTQISAGDYKSSINDAAGRANNDATRVIVNVSPTALTR	518
Qy	527	SVSRSGIESQGRSADASORAAETIVRDSQTLGDVYSRLQVLDLSMTSTIVSNPOAQEEIM	586
Dd	519	SVPRKRTLEARG-PEKTTDALARVIVSGNSRTLDGVYSQVSALDSVQIITQSNQANNEIR	577
Qy	567	OKLTASISKAPQFGYPAYVONSADSLQKFAQALEREFPVDGERSLAESQENAFRQAPATQO	646
Dd	578	OKLTASVATKPPQFGYPYVQLSNDSTQKFTAKLESFAEGSRFAETIKALSFETNSLFTQO	637
Qy	647	VLVNATSLFSGYL 659	
Dd	638	VLVNIQSLSYGL 650	
RESULT 11			
ABB94277			
ID	ABB94277	standard; Protein: 651 AA.	
XX	AC		
XX	ABB94277;		
XX	DT	05-JUN-2002 (first entry)	
XX	DE	Chlamydia pneumoniae protein sequence SFQ ID NO:405.	
XX	KW	Chlamydia infection; Chlamydia; Vaccine; detection; diagnosis;	
XX	KW	antigen; antibacterial; immunostimulant; immune response;	
XX	KW	Chlamydia-specific T-cell response.	
XX	OS	Chlamydia pneumoniae.	
XX	PN	W0200208267-A2.	
XX	PD	31-JAN-2002.	
XX	PE	20-JUL-2001; 2001WO-US23121.	
XX	PR	20-JUL-2000; 2000US-0620412.	
XX	PR	23-APR-2001; 2001US-0841132.	
XX	PA	(CORI-) CORIXA CORP.	
XX	PI	Fling SP, Skeiky YAM, Probst P, Bhatia A;	
XX	DR	WPI: 2002-179901/73.	
XX	PT	Novel compositions comprising Chlamydia Cap1 protein and its use in the	
XX	XX	treatment of Chlamydia infection -	

Disclosure: Page 370-372; 537pp: English.

The present invention describes compositions comprising a Chlamydia CapI protein and methods for the diagnosis and therapy of Chlamydia infection. Chlamydia DNA and protein sequences from the present invention can have antibacterial and immunostimulant activities, and can be used in vaccines. Compounds from the present invention can be used for eliciting an immune response, specifically stimulating a Chlamydia-specific T-cell response or inhibiting the development of a Chlamydia infection in an animal. Methods from the present invention can be used: for detecting the presence of Chlamydia in a patient; to stimulate and/or expand T cells specific for a Chlamydia protein; and for treatment of a Chlamydia infection. ABL92394 to ABL92709 and ABB94096 to ABB94374 represent sequences used in the exemplification of the present invention.

Sequence 651 AA:

Query Match 27.9%; Score 894; DB 23; Length 651;
Best Local Similarity 34.6%; Pred. No. 8,4e-46;
Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

QY 11 GPESYSSNOSSNNPIINQIASNSPTKSTKA-----SEASP--SASSVSWSPLSSAK 63
DB 6 GGPPIDETERPPADLSAQGLEASAAASAEQRTAGAEAKPKESKTQSVBERWSTLNRSAV 65
QY 64 NALISLRD--AILNKNSPTDSLSOLEASTSTVTYRAAADYDEAKSNPPTASGLEENA 121
DB 66 NALMSLRADKLGIASNSSTSR--ADVDSMTATPTPTPTPDYKQAQTAFTITTS 124
QY 122 KTLAEFTKMDLMAALODMERLANSDESNHTEEVNNIKALEAKQDTID-----K 173
DB 125 TSLADIQALVSLDPAVNIKDTAATD-----EET-----AIAAEWETKNADAVKGAQ 173
QY 174 LNKIVTTLONOKSLFEVLTKTDSAQQIPAINSQLEINKNSAQIILKDLERQNI---SVEA 230
DB 174 ITELKAYASDQALIIDISGKLTSPDLQALILQSYANNNNKALELKEQNDPVPYCKTPA 233
QY 231 VLTNAGEVYKASSESGILGQALOSIVAGDQSQAAVYLAQAOQNSNPDIMATKELIDAE 290
DB 234 IAQSLVDDTDTATPQIEKDGNAIRAYFAGNAGASAVENAKSNNSISNIDSKAMAIATAK 293
QY 291 TKVNELKOEHTGLDPLVYKKAEOISQAQKDIOEIRPS--GSDIPYGPS--GSAASAGSA 348
DB 294 TQIAEAOCK---FPDSPITQEAQWVYQAEKDKNIKIPADSDGVPNPGTIVGSGKQCGSS 350
QY 349 AGALKSSNNSGRIILLDDVDNEMAIALOQFRSMLIEDFNNPNPTAKELQAMEKQLTAM 408
DB 351 TGS1-----RVSKLLDDEAENETASILMSGROMIHFNENPNPSQAQOGLAQAARAA 403
QY 409 SDOLVAGADGELPAETQAIKDALQAL--KOPSADGLATMGQVAFAPAAVGVGSGAGTAGTV 467
DB 404 K---AAGDSSAAALADQAKALEALGKAGQOQGLNMLGQIASAAVYSAGVPPRAASTI 460
QY 468 QMNVKQLYKTAFFSTSSSSYAAALSDGYSAYKTILNSLYSESR--GVQASISOTANPALSR 526
DB 461 GSSVQILKYKT--SKTSGDYKTQISAGYDAVKXINDAYGARANDATRDVINNVSPATLPR 518
QY 527 SVSRGEGISQGRSADASQRAETIYRDSQDTGLDYSYRQVLDLSMTSTIVSNPQANQOEIM 586
DB 519 SVPRARTFARG--PEKTDQALARVIGNSERTLGDVYSQVSAQSQWQIIQSNPQANNEIR 577
QY 587 OKLTASISKAPOFGPAYQNASDSLOKRAAOLEREFEVNGERSLAEQENAPKQAPFIOQ 646
DB 578 OKLTASVATKPPQFGPIYVQLSDSTOKETAKLESLEAEGSKRTAEIKALSFTETNSLFIQ 637
QY 647 VLVNIASLFGSYL 659
DB 638 VLVNIGSLXSYL 650

RESULT 12
AAV71955
ID AAV71955 standard: Protein: 583 AA

```

XX AAY71955;
AC 26-MAR-2001 (first entry)
XX
XX 5'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
XX 76 kDa protein: bactericidal; diagnosis: prevention;
XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX acute respiratory disease; cough; sore throat; hoarseness; fever;
XX vaccine; immunisation; treatment; truncation mutant; mutin.
XX
XX Chlamydia pneumoniae.
XX Synthetic.
XX
XX Key Location/Qualifiers
XX Misc-difference 497 /note= "Encoded by ACT"
XX
XX WO200066739-A2.
XX
XX 09-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-CA00511.
XX
XX 03-MAY-1999; 99US-01332270.
XX 30-JUN-1999; 99US-0141276.
XX
XX (AVENTIS PASTEUR LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX
XX WPI: 2000-687542/67.
XX N-PSDB: AAD02064.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 16b: Page 100-102; 112pp: English.
XX
XX
XX The present sequence is 5'-truncated
XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
XX protein is used in the
XX diagnosis, prevention and treatment of C. pneumoniae
XX infections (e.g. pneumonia, upper respiratory
XX tract disease, bronchitis, sinusitis and acute respiratory
XX disease such as cough, sore throat, hoarseness, fever; and
XX abnormal chest sounds on auscultation). C. pneumoniae sequence
XX is also used as vaccines for immunising humans against diseases
XX caused by C. pneumoniae.
XX
XX Sequence 583 AA:
SQ
Query Match 25.9%; Score 830.5; DB 21; Length 583;
Best Local Similarity 34.9%; Pred. No. 4.9e-42;
Matches 213; Conservative 117; Mismatches 235; Indels 45; Gaps 15;
OY 67 ISLRD--AIIKNNSSPTDSLSQLEASTSTSTVTRAAKDYDEAKSNFTAKSGLENAKTL 124
DB 1 MSLADKLGIASSNSSSSSTRS-ADVSTTATATPTPPPTFDDYKTKQAQVADPTIFSTSL 59
OY 125 AEYETKRMADLMAADMELANSDPSNNITEVNNIKKALEOKTID-----KLKK 176
DB 60 ADIQALVSLQDAVNIKTKTAATD-----EET-----AIAEWEKKNDAVKGAGQITE 108
OY 177 LVTLQNNKSLTEVLKTTSDADQIPAINSQLKINKNSADQIKKDERQNI---SYEAVLT 233
DB 109 LAKYASDNQAIIDSLCKLTPDLQALQSVANNKKALEKMDNPNVVGKTPAIHQ 168
OY 234 NGEVYKASSEAGIKLQALQSTVDAGDOSQAAYLQAOQNNSPDNIAATKELIDAAETKV 293
DB 169 SLVDQTDATATQIEKDGNAIRDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATFAKTOI 228

```

```

OY 294 NELKOEHTGLTDSPLVKKAEQISOAKDQIEIKPS-GSDIPVGP-CSAASAGSAGA 351
DB 229 AFAQKK---FPDSPILQEBQWVIOAEKDLKIKRADSDVNPRTTYGSGKQOGSSIGS 285
OY 352 LKSSNNGRIISLLDDVDNEMAAIALQGRSMIEQFNWNNPATKELQAMEAQLTAMSQ 411
DB 286 I-----RVSMILDDAEHETASILMSGFRQIHMFNTEPNPSQAQOELAQARAAK-- 336
OY 412 LVGAGELPAELQAIKDALQAL-KQPSADGIATAMGOVAFPAAKVGGSGAGTACTYQMN 470
DB 337 -AAGDSSAAALADLQKALEALGRAGQOQGLNMLQJASAAVVSAGVPAASSIGSS 395
OY 471 VKQLYKTAFTSSSSSYAALSDGYSAYKTLNLSYSESPS-GVQSAISOTANPALSRVS 529
DB 396 VKQLYKT--SKSTGSDYKQISAGYDAYKSIDATGRANDTRDVYNNVSPALFTRSP 453
OY 530 RSGIESQGRSADASQRAETIVRDSQTLGDVYSRLQVLDLSMTSTVSNPQAOEIMOKL 589
DB 454 RARTEARG-PEKTDQALARVIGNSRTLDVYSQVSAQSVQIIOINPQANNEEIROKL 512
OY 590 TASISKAPQFGYPAYQNSADSIQKFAQLEREFVNGERSLASQENAPRKOPAFIOQVLY 649
DB 513 TSAVTKPPQFGYPVQLSNDSTOKFIKLESIFAEGSRTAECIKALSFETNSLFOOVLY 572
OY 650 NIASLFSGYL 659
DB 573 NIGSLXGYL 582

```

```

RESULT 13
AAY71956
ID AAY71956 standard; Protein; 452 AA.
XX
XX AAY71956;
XX
XX 26-MAR-2001 (first entry)
XX
XX 3'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
XX 76 kDa protein: bactericidal; diagnosis: prevention;
XX pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX acute respiratory disease; cough; sore throat; hoarseness; fever;
XX vaccine; immunisation; treatment; truncation mutant; mutin.
XX
XX Chlamydia pneumoniae.
XX Synthetic.
XX
XX WO200066739-A2.
XX
XX 09-NOV-2000.
XX
XX 03-MAY-2000; 2000WO-CA00511.
XX
XX 03-MAY-1999; 99US-01332270.
XX 30-JUN-1999; 99US-0141276.
XX
XX (AVENTIS PASTEUR LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
XX
XX WPI: 2000-687542/67.
XX N-PSDB: AAD02065.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 16c: Page 104-106; 112pp: English.
XX
XX The present sequence is 3'-truncated
XX Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
XX protein is used in the
XX diagnosis, prevention and treatment of C. pneumoniae
XX infections (e.g. pneumonia, upper respiratory

```


DT	13-SEP-1999	(first entry)	
XX			
DE	Chlamydia pneumoniae	surface exposed polypeptide.	
KW	Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;		
KW	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;		
XX	vaccine; neutralising epitope.		
XX			
OS	Chlamydia pneumoniae.		
PN	MO92927105-A2.		
XX			
PD	03-JUN-1999.		
XX			
PF	20-NOV-1998;	98WO-1B01890.	
XX			
PR	04-NOV-1998;	98US-0107078.	
XX	21-NOV-1997;	97FR-0014673.	
PA	(CEST) GENSET.		
XX			
PI	Griffais R;		
XX			
DR	WPI: 1999-357842/30.		
XX			
PT	Genome sequence of Chlamydia pneumoniae		
XX			
RS	Page 1156; Disclosure; 1912pp; English.		
CC	AAV34584-Y35879 represent the proteins encoded by all the open reading		
CC	frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.		
CC	C. pneumoniae causes respiratory disease such as pneumonia and		
CC	bronchitis and is thought to be a contributing factor in heart		
CC	disease, sarcoidosis, sinusitis, purulent otitis media, erythema		
CC	nodosum or pharyngitis. The polypeptides encoded by the open reading		
CC	frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in		
CC	immunogenic compositions as vaccines. Vectors containing C. pneumoniae		
CC	nucleotide sequences can also be used as immunogenic compositions,		
CC	especially where the vector directs the expression of a neutralising		
CC	epitope of C. pneumoniae.		
XX			
50	Sequence	478 AA:	
	Query Match	14.1%; Score 451; DB 20; Length 478;	
	Best Local Similarity	29.0%; Pred. No. 2.9e-19;	
	Matches 135; Conservative	94; Mismatches 189; Indels 48; Gaps	14;
OY	11 GPEVSNSNOSSMNPINQOIASNSETKSTKA-----SEASP--SASSSVSSWSFLSSAK	63	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		
DB	8 GGPIDETFRTPADLSAQGLEASANKSAEQRITAGAEAKRKESKTDSEVWLSRAV	67	
OY	64 NALJLRD--ALNNKNSPTDSLSOLEASTSTSYTRAAADYDDAKSNFTAKSGLENA	121	
	: : :		
DB	68 NALMSLAKLGLASSNSSTSR--ADVDTATATPTPPPTFDYKTOAQAYDITFTS	126	
OY	122 KTLAEYETKMDLMAALDOMERLANSDDSNHTEEVNNIKKLEAKOKPTID-----K	173	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		
DB	127 TSLAIQALVSLQDAVNRINKDTATD-----EET-----ATAEMETKKNDAVKGAQ	175	
OY	174 LNKLTVLONONKSLTEVLKTTDSADQIPALNSOLEINKNSADQITKDLERONI---SYEA	230	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		
DB	176 ITELAKVYASDNQAILIDISGKLTSPFLLOALLQSVAAANNKKAHELKEMQDNVYVGRTPA	235	
OY	231 VITNNGEYIKASSEGGIKGALQSLIVDAGDSQAAYVLAQOQNSPDPNIATATKELIDAE	290	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		
DB	236 IAOQSLVDDTDAATATIEKDGNAIRAYFAGNASCVAVEAKKSNNSISNDSAKAAIATAK	295	
OY	291 TKVNEKQEHQHTLQDSPLCVKKAEQISOAKODIOETKPS--GSDPIVGPS--GSAASAGSA	348	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		
DB	296 TQIALAQKK---FDPSPILQEAQWYIQAEKULKNIKPADGSDVNPPTGVYGGSQGSS	352	
OY	349 AGALSSNNSGRIISLLDDVDEMAAIALQGRSMIEQFNVNNPATAKELQAMEAQLTAM	408	
	: : : : : : : : : : : : : : : : : : : : : : : : : :		

QY	DB	Sequence	Score	DB	Length	Indels	Gaps
QY	353	IGSI-----RVSLMLDDEAENETASIIISGFRQHIHFNTENPDUSQAQOELAAQARAA	44.5	20	212	7	4
QY	409	SDQLVGADGELPAEIQAIKDALAQL-KOPSADGLATAMGOVAFPA	45.3	20	212	7	4
DB	406	K---AAGCDSAAMAALADAKKLEALGKAGGQGGITLNLGQIASAA	44.8	20	212	7	4
QY	16	RESULT					
ID	AAV35357	standard; Protein: 212 AA.					
QY	AAV35357						
QY	AAV35357						
QY	13-SEP-1999	(first entry)					
QY	Chlamydia pneumoniae	transmembrane protein sequence.					
QY	Respiratory disease: pneumonia; bronchitis; heart disease; sarcoidosis;						
QY	sinusitis; purulent otitis media; erythema nodosum; pharyngitis;						
QY	vaccine; neutralising epitope.						
QY	Chlamydia pneumoniae.						
QY	MO9927105-A2.						
QY	03-JUN-1999.						
QY	20-NOV-1998;	98WO-IB01890.					
QY	04-NOV-1998;	98US-0107078.					
QY	21-NOV-1997;	97FR-0014673.					
QY	(GEST) GENSET.						
QY	Griffats R;						
QY	WPI, 1999-357842/30.						
QY	Genome sequence of Chlamydia pneumoniae						
QY	Page 1155; Disclosure; 1912pp; English.						
QY	AAV34584-Y35879	represent the proteins encoded by all the open reading frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.					
QY	C. pneumoniae	causes respiratory disease such as pneumonia and bronchitis and is thought to be a contributing factor in heart disease, sarcoidosis, sinusitis, purulent otitis media, erythema nodosum or pharyngitis. The polypeptides encoded by the open reading frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in immunogenic compositions as vaccines. Vectors containing C. pneumoniae nucleotide sequences can also be used as immunogenic compositions, especially where the vector directs the expression of a neutralising epitope of C. pneumoniae.					
QY	Sequence	212 AA;					
QY	Query Match	13.9%; Score 445.5; DB 20; Length 212;					
QY	Best Local Similarity	47.6%; Fred. No. 2.1e-19;					
QY	Matches 100; Conservative	34; Mismatches 69; Indels 7; Gaps 4;					
QY	451	FAAAVGGGAGTACTGTCVGMNKKYKTAFFSSSTSSSYAALSDGYSAYKTLNLSSESRS	510				
QY	8	FVSAGVPPRAAASSTIS---SVNQLYKT--SKSTGSDYRTQISAGVDAYKSTINDAYGRARN	62				
QY	511	-GVASISOTANPALSRVSRSRGIESQGRSADASQRAAETIVRDSOTGLDVSRLQVLDS	569				
QY	63	DATRDVINNVSPALTRSPARFARCG-PEKTDALARVISGNSRTIGDVSQVVALOS	121				
QY	570	LMSTIVSNPQANQDEIMKRTASISKAQOFQPRAYONASDSLQKRAALDERFVNGERSL	629				
QY	122	WQIIQSNQDQANNEIRKRLTSAVTKPPQFGPYVALSNDSTOKFIATLESLEFAGSRFA	181				
QY	630	ASQENAFKPKAPFIOQVLVNIASFSSYL	659				

Db 182 AETKALSFETNSLFIQQVLVNLISLXGYL 211
RESULT 17
AAU37120
ID AAU37120 standard; Protein; 2344 AA.
XX AAU37120;
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #1290.
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
PN WO200170955-A2.
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 22-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HR;
XX
XX WPI: 2001-611495/70.
DR N-PSDB; AAS54979.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12713; 511pp; English.
PS
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2344 AA:
SQ

Query Match 8.0%; Score 256; DB 22; Length 2344;
Best Local Similarity 21.1%; Pred. No. 1,4e-06;
Matches 148; Conservative 124; Mismatches 307; Indels 124; Gaps 21;

Oy 9 ESGPESVSSNOSMNPINQAIASNETKSTKASAPSPASSSVSWFSLSSAKNALIS 68

Db 1177 DSISEAISGSESTISLSSESNSTDSSEKSAFALESESESTSESLSGSTDSST 1236
Oy 69 LRDPAIINKNSPPTDSLQLEASTSTVTFRVAAKDYEAKSNFDTAKSGLENKATXEYE 128
Db 1237 LSDSNSESGSTSTSLSNSTSGSASISTST-----SGSASTYKAS--ESVSTLSST 1286
Oy 129 TKMADLMAALQMERLANSPPSNHTEEVNNIKKALEAKDITDKLNKLVTLQNMKSLT 188
Db 1287 TS-----TSLDSTSLSTSLDSTSGSKSN---SLSAKMTSDSIS---TRKSESLAS 1334
Oy 189 EVLKTDSADQIPAINSOLETKNNSADQITKLEFRONISYEAVLTNAGEYIKASAGIK 248
Db 1335 TSLSGSTSESESGSTSS---SESKSDSTSMLSMSQSTSGSTSVSTSESLDSTSTSL 1390
Oy 249 LGOAL-OSIYDAGDOAAVLQAQNNPNINATKELIDPAATKYNELKQETGLTDP 307
Db 1391 LSAAMQSGVDSNASASASTSTSTSESDSOSTSS-YTSQSTSGSESTSTSLDST 1449
Oy 308 LVKKAEEQI-----SOAKDQIEIKPSG-----SDIPVPGSGSA-- 342
Db 1450 SISKSTSGSGSTSTSAISLSESESDQSTSTSTSEKSESTSTSLDSTSTNSGSAST 1509
Oy 343 -----ASNGSAAGALKSSNNNGRISLLIDVDNEMAIALOGFMSIEQFVNNPAT 395
Db 1510 STLLSNASASESDSSTSLDSTSTASMSQSESDSOSTSTSL-----NSQSTST 1560
Oy 396 KELQAEAOQLTAMSDQVLVAGDGLPAEIOAIKALAOA---LKQPSADGLT----- 444
Db 1561 IRNSTIASSESVSESTSESGSTSESTSESDSTSLDSTSTSNSTASGSASTSTSD 1620
Oy 445 -----AMGVAPAAKVGCG-----SAGTAGTVQNM 470
Db 1621 RSTASSTSMRTSLDQSMSTSTSTSVSDSTSLDSTSDSTSDSTSTSTSGMSAS 1680
Oy 471 VKQLYTAPSTSSSYAAALSDGYSAKYTLNLVYESNGVSAISQTPANPLSSVSR 530
Db 1681 ISLSDSTSTSTASAEVWSASISDSQMSBESVNDSESVSESNSE---SDSKSMGSTSVSD 1737
Oy 531 SGIESQGRSADASQRAAETI-VRDSOTLGDVYRQLQVLDSLMSTIVSNPOANEETIMOK 589
Db 1738 SGLSVSTSLRKSESVSESLSGSQMSDVS---TSDS-SLSVSTSGRSESVSESD 1793
Oy 590 TASISKAPQFGYPAYONSADSLQKFAOLEREFDERSIAES 632
Db 1794 SLSDSKSTS-GSTSTSTSG-SLSTSTSLSGSESVSESSLSDS 1834
RESULT 18
AAU37403
ID AAU37403 standard; Protein; 6281 AA.
XX
XX AAU37403;
XX
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #1573.
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
PN WO200170955-A2.
PD 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.

CC a sequence present in *Lactobacillus rhamnosus* strain HN001 and encoding a
 CC polypeptide capable of modifying the flavour, aroma, texture, nutritional
 CC and health benefits of milk-derived products, and/or survivability of
 CC microbes in dairy manufacturing processes. The polynucleotides are useful
 CC for improving the properties of microbes used in the manufacture of milk-
 CC derived products such as cheeses, yogurt, fermented milk products, sour
 CC milks and buttermilk; in modifying the flavour, aroma, texture and health
 CC -related benefits of milk-derived products and in increasing the survival
 CC of microbes during industrial fermentation processes. The bacteria may be
 CC used to increase resistance to enteric pathogens and anti-infection
 CC activity, including treatment of rotavirus infection and infantile
 CC diarrhoea, aid in lactose digestion; as anti-cancer and anti-mutagenesis;
 CC liver cancer reduction; reduction of small bowel bacterial overgrowth;
 CC immune system modulation and treatment of autoimmune disorders and
 CC allergies; treatment of allergic responses to foods; reduction of blood
 CC lipids and prevention of heart disease; antihypertensive effect;
 CC prevention and treatment of urogenital infections, *Helicobacter pylori*,
 CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and
 CC irritable bowel syndrome; modulation of endocarditis; and for improved
 CC protein and carbohydrate utilization and conversion. The transgenic
 CC microbial population can be administered to a mammal as an anti-
 CC carcinogenic agent. The present sequence is *Lactobacillus rhamnosus*
 CC outer membrane protein rompa.

XX Sequence 1463 AA;

Query Match 7.8%; Score 251; DB 23; Length 1463;

Best Local Similarity 25.2%; Pred. No. 1.5e-06;
 Matches 165; Conservative 82; Mismatches 295; Indels 114; Gaps 26;

QY 4 HHHHMEGPEVSSNOSSNMPTIINGQIASNSETKSEASPSASSVSSWSFLSAK 63
 DB 751 HADNMEI--KSLASDAEKQSOI--ALAAKSAASASSAAAIIVASSASE---ASSA 802
 QY 64 NALISRDALILKNSPTPTSLSEASTSTYTRVAAD--YDPAKSNFTAKGLEN 120
 DB 803 AAASVASDAASANSAAAYDYSTAS--EASAS--AANDSGYATAPASAAASAAAMSA 855
 QY 121 AKTLAEYERKMA--DLMAA--LQDMERLANSDPNSNHTTEVNNIKKALEAKOTIDILNK 176
 DB 856 ALSTAOVAAKVAASDAALAAASAAVAASAAQSSKKNQA-----TAATRSQALDLNK 908
 QY 177 LVTLOQNMKSLTEVLEKTTSDADQIPAINSOLEINKNSADQIIKLERONISYEAVLTIMG 236
 DB 909 I-----KSLFDYASGASS-----SASEGQAST-----ATSAVAS 938
 QY 237 EYTKASSEGKILQALQSTVDA---GPOSQAAYVQAOQNNSPDIAATKELIDAEIK 292
 DB 939 AASSASSEAGSYAHQAGSSASDAVGGSGAQAHAATAAASSTYPRDSGIOTSLASQAASE 998
 QY 293 VVELKQEHGLTDSPLVKKAEQISOAKDIOEIKPSGSDIPYVGPSSAASGSAAGAL 352
 DB 999 AAKASSNAASATSAAVG-----FSAASDASBOAKTAASADYVA--SSAATSTANSASAA 1051
 QY 353 KSSNNSGRISLLIDVDYEMAAITALQGERSMIEQFNVNNPATAKELQAEQOLTAMSQDL 412
 DB 1052 ASATKAG-----DSKAAA-----GFSSAASA--AASSAKGAEAVASEAASAAASDDS 1096
 QY 413 VGADGELPAEIQAIKALQALQKOPADGLATAMGOVAPAAKVGGS-----AGTAGIV 467
 DB 1097 VASSA---ASAAAGFDKASA--AEGNAASASAAASAAAGCTRGASSASASEAGAST-- 1151
 QY 468 QMNVKOLYTAFTSTSSSSSYAAALSDGYAYKTLINLSESRSGVQSASIQ--TANPALS 525
 DB 1152 ----ATSVYASAASSSSSEAGSYAHQAGSSA-----SEATGHASASISQASAAASAS 1200
 QY 526 RSVSRGIESQ-----GRSADASQRAAETIVRDSQTLGDVYSRLQVLDLSMT--IYVNPQ 579
 DB 1201 RVPDSGIGSDVSIASSASTASTASASAAQSEASTASTASASHASEQASTASSSDVYSSA 1260
 QY 580 ANQELMQLTISAKAPQFGYPAVQNSADSLQKFAQLEREVDGERSLAEQEN 635
 DB 1261 ASVASSAASAAASAAKAGNSAAGTYSHAASAAASAKGAESQASASAAASAAASDDS 1316

RESULT 20
 AAG82935
 ID AAG82935 standard; Protein; 5024 AA.

AC AAG82935;
 AC AAG82935;
 DT 03-SEP-2001 (first entry)

DE S. epidermidis open reading frame protein sequence SEQ ID NO:2964.

KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.

OS Staphylococcus epidermidis.

PN WO200134809-A2.

PD 17-MAY-2001.

PF 09-NOV-2000; 2000WO-US30782.

PR 09-NOV-1999; 99US-0164258.

PA (GLAXO) GLAXO GROUP LTD.

PI Kimmery W;

PR WPI: 2001-316495/33.

DR N-PSDB; AAH53785.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,
 useful for vaccinating against infections, e.g. endocarditis -

PS Claim 18; Page 779-781; 2168bp; English.

CC AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce host cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH55091 to
 CC AAH55098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.

XX Sequence 5024 AA;

Query Match 7.8%; Score 249.5; DB 22; Length 5024;

Best Local Similarity 19.8%; Pred. No. 9.1e-06;
 Matches 148; Conservative 130; Mismatches 247; Indels 221; Gaps 33;

QY 1 MHHHHHMEGPEVSSNOSSNMPTIINGQIASNSETKSEASPSASSVSSWSFLS 60
 DB 1042 VNNAGQHVIDEQATLSSD--TTNQLANAYTQAKSNLHGDTKLQHKDKSAKOTIAQLQNLN 1099
 QY 61 SAKNALISRDALILKNSPTD---SLSOLEASTSTSTYTRVAARD----- 103
 DB 1100 SAKC---IHEDSLIDWSESTRVOYQHDLEAQLDGMGLAKESIKDNTRYIVSNGNTYINAE 1156
 QY 104 -----YDEA-----KSNPDTAKSGLENAKTLAEYETKMADLMAALAD 140

```
Db 1157 PSKQAYDAAYVQNAQNIINGTNOPTINKGNVTATQTVAKTRKD-----ALDGC 1203
Oy 141 MERLANSPPSNHTEEVNNIKKALEAKD-----TIRKLNK-LVTLONONKSLTE 189
Db 1204 DHRLEAKKNNANQITIRLSMNNNAQDAEKNLVNASASTLEQVOONLOTAOOLDNAMKE 1261
Oy 190 VLKTTDSADQIPA-----INSOLEINKNSADQIITKDERONISYEAVLTNAGEVIRASSDA 245
Db 1262 LRQSTANKQDVAKADSKYLNEDPQIKONVDDAY-----ORVETIINETONELUKAN--- 1312
Oy 246 GIKLQALOSTIVDAGDSQAAYLQAQONNSPDNIATK-----ELIDAEKTVNE----- 295
Db 1313 ---IDQANQSVQNA-----EQALHGAEKLNQDKQSTSTELDGLTDLTDAQREKLREQINTS 1365
Oy 296 -----LQKQ-----EHTGLTDSPLYKKAEEQISQAKQIDQETKPGSDIPYVPGSGSAASACS 347
Db 1366 NSRDDIKQKIEQAKALNDA--MKKLKEQV--AQKDGVBH---ANSDY-----TNEDSAQKDA 1414
Oy 348 AAGALKSS-----NSGRISLLDDVDNEMAALA-----LQGRSMIEQFNVPNNPATAKEL 398
Db 1415 YNNALKQAEADITNNSSNPFLNADITNLNLIKQADNLHGAQKLDQDKNTTNOAIGNLN 1474
Oy 399 QAMEAQLTAMSDQLVGADGELPAETQAIKDALQALQKPSADGLATAMGOVAFMAAKYVG 458
Db 1475 HLNQPOKDALIQAINQATS-----RDQVAKLKE--AEALDEAM----- 1511
Oy 459 GSAGTAGIVQNVNKKYKTAFASTSSSYAALSDQISAKYKTLNLSYSGSGVQSAISQ 518
Db 1512 -----KOLEDOVNODDQISNSPFINESDQKQ-----YNNKIQAAKEIINO 1554
Oy 519 TANPALSRVSRSIGESQGRSADASORAETI--VRD-----SQTLDGVYSRLQ 565
Db 1555 TSNPTLDR-----OKIDTLONIKDAVNNLHGDKLQAKQSDANNOQLN 1597
Oy 566 VLDSI-----MSTIVSNPOANOE-----EIMQKLTASISKAPQFGYPAYONSADSLQK 613
Db 1598 HLDLITEQKNHFPLINNADTRDEVNQKLEIAKQLDNGDMSTLHK-----VINDKQIQIH 1652
Oy 614 FAQOLEREFVQGERSLAESQENAFK 639
Db 1653 LS-----NYINADNDKKQNYDNAIKE 1673

RESULT 21
ABP38314
ID ABP38314 standard; Protein; 10182 AA.
AC
XX ABP38314;
XX
XX 24-JUL-2002 (first entry)
DT
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:3159.
DE
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW
XX Staphylococcus epidermidis;
XX
XX Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-0134001.
XX
XX 14-AUG-1997; 97US-055779P.
XX
XX 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucelte-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
```

```
DR N-PSDB; ABN90859.
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PS polypeptide, useful for diagnosing and treating bacterial infections -
XX
XX Disclosure; SEQ ID 3159; 267pp; English.
XX
XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
XX Sequence 10182 AA:
SQ
Query Match 7.8%; Score 248.5; DB 23; Length 10182;
Best Local Similarity 19.7%; Pred. No. 2.7e-05;
Matches 141; Conservative 128; Mismatches 256; Indels 189; Gaps 31;
Oy 18 NSSMNPITINGQIASNETKESTKASEASPSASVSSWSPFLSARKNLISRLAIILKN 77
Db 6003 SSOTINQNLNNAVTOAKSNLHGDTKLQHKDSKAKOTIQLONLNSAKQ---IHEDSLIDNE 6059
Oy 78 SSPTD---SLQLEASTSTSTVTVAANDYDEAKSN-----FDTAKGLENA 121
Db 6060 SYRTQVQHDLTBAQLDLGMLGALKESIKDYTNIVSGNYINAEPKSKQAYDAAYVQNAQNI 6119
Oy 122 KTLAEYET--KAMDLMALADMERLANS-----DPSNNHTEEVNNIKKALEAKD--- 169
Db 6120 INGTNOPTINKGNVTATQTVAKTRKDALDGDHRLREAKNNANQITIRLSLNNAQKDAEK 6179
Oy 170 -----TIDKLNK-LVTLONONKSLTEVKTDSADQIPA-----INSOLEINKNSADQI 217
Db 6180 NLVNASSTLEQVQNLQTLAQOLDNNAMGELRQSIYAKDKQKADSKLTNEBPQIKONTDDAY 6239
Oy 218 INDLERONISYEAVLTNAGEVIRKASSEAGIKLQALQSLIVDAGDSQAAYVLAQONNSPD 277
Db 6240 -----QVETIINETQNELKAN-----IDQATQSVQNA-----EQALHGAEKLNQK 6283
Oy 278 NTAATK-----ELIDAEKTVNE-----LQKQ-----EHTGLTDSPLYKKAEEQISQA 319
Db 6284 QTSSTELDGLTDLTDAQREKLREQINTSNSRDDIKQIEQAKALNDA--MKKLKEQV--A 6339
Oy 320 OKDIEIKPSSGSDIPYVPGSGSAASAGSAGALKSS-----NSGRISLLDDVDNEMAAL 375
Db 6340 QKDGVBH---ANSDY-----TNEDSAOKDAYNNALKQAEADITNNSSNPFLNADITNLNNTI 6392
Oy 376 A-----LQGRSMIEQFNVPNNPATAKELQAMEAQLTAMSDQLVGADGELPAETQAIKDAL 430
Db 6393 KQADONLHGAQKLDQDKNTTNOAIGNLNHLNQPOKDALIQAINQATS-----RDQV 6443
Oy 431 AQALQKPSADGLATAMGOVAFMAAKYVGGSACTAGTVQNVQLYKTAFASTSSSYAAA 490
Db 6444 AEKLEKE--AEALDEAM-----KQLEDOVNODDQISNSSPF 6476
Oy 491 LSDGSAYVTLNLSYSESRSQVQSAISOTANPALSRVSRGIESQGRSADASORAETI 550
Db 6477 INEDSDQKQK---YNDKIQAAKEIINQTSNPTLQK-----OKIADTL 6515
Oy 551 ---VRD-----SQTLDGVYSRLQVLDL-----MSTIVSNPOANOE-----EI 585
Db 6516 QNIKDAVNNLHGDKLQAKQSDANNOQLNHLDDLTQEOKNHFRPLINNADTRDEVNQKLEI 6575
Oy 586 MOKLTASISKAPQFGYPAYONSADSLQKFAQLEPFVQGERSLAESQENAFK 639
Db 6576 AKQLDNGDMSTLHK-----VINDKQIQIHLS-----NYINADNDKKQNYDNAIKE 6619
```


DR N-PSDB; AAS52198.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5835; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2434 AA;

Query Match 7.5%; Score 239.5; DB 22; Length 2434;
Best Local Similarity 21.4%; Pred. No. 1.4e-05;
Matches 147; Conservative 107; Mismatches 222; Indels 211; Gaps 31;

QY 21 SNMPTINGOIASNETKESTKASESPSASSSVSMFLSKAKNALISRDATLNKNSP 80
DB 1889 AMERLING-IQKDKQYKQSVNFTDADPEK-----QTAYNAVTAENITINQANGT 1937
QY 81 TDSLQLEASTSTVTVRAA---KDYDEAKSNFTAKSGLEN----- 120
DB 1938 NANOQVEAALSTVTTKQALNGDKRVTDKANNANQTLSTLDLNAQGAVTGNINQAH 1997
QY 121 -----AKTAEYETKADLMALQD-----MERLANSBP--SNHTEEVNITKAL 164
DB 1998 TVAEVTQAIQTAEQELMTAGNLKNSLNDKDTLGSQNFADADPEKKNAVNAEHNMENLT 2057
QY 165 EAQKDTIDKLKLVTLQONKSL-TEVLKTPDSADQIPAINSOLEINK---NSADQIID 220
DB 2058 -----NKSTGTNPK-----DOVEAAMQVNAITKALNCTOMLEKA 2093
QY 221 LERONISYEAV--LTNAGEVIRKASSEAGIKGOALQSTVDAGDQGAVALQAQNNSPDN 278
DB 2094 KOHANTAIIGLSHTWNAQK-----EALKQIV-----QGSTVAAEQNGQOKAN 2136
QY 279 IATKELIDAAETKV-----NELKQEHGTLDSPLYKR--AEQJISQAQKIDQETKPSG 330
DB 2137 -----NVDAAADKIKQSIADNATTKQONQNYTDAQONKDAYNNAVTTOGIIIDTTSPT 2190
QY 331 SDPIYVPGSASAGSAGALKSSNNSGRISLLDDVQENMAALQGFERSMIEGFVNN 390
DB 2191 LDPTVY-----NOAAQGVSTTKNA-----LNGNEN-LEAAKQQAQSSQISGLSDNIN 2234
QY 391 NATAKELQAMBAQJLTAMSQDLVGADGELPAETQAIKDALQALQPSADGLATANGOVA 450
DB 2235 N-----AQKQTVTDQINGA-----HTWDE--ANQIKQ-NMQNLNTANG---- 2269
QY 451 FFAAAKVGGSAGTAGTQVQNNKQIKTAFSSSTSSSYAALSDGSAVYTLNLSYSESS 510
DB 2270 -----NLKQ-----AIDKDKATKATVNF-----T 2288
QY 511 GVQSAISQTPANPALSRVSRSIGESGRSADAQRAAETIVD-----SGTLGDVYSRLQV 566
DB 2289 DADQAKQQAQYNTAVTWNA---ENITISKANGCNATQAEVEQAIKOVNAKQALNGANVQHA 2345

QY 567 LDSLMSTIVSN---POANOEIMQKLTASISKAPQFGYPAYONSADSLQKFAQJEREFV 623
DB 2346 KDEATLINSNDLNOAQKDALQGVQNAVTVA---GVNNVKTQAEQLNNMNTQKQGI 2402
QY 624 DGERSLAE-----SOENAFRKQPA 642
DB 2403 DKEQTKADGNEFVNADPDKOMAYNQAYA 2429

RESULT 25
ABP39618
ID ABP39618 standard; Protein; 2137 AA.
XX
XX ABP39618;
AC
XX
XX 24-JUL-2002 (first entry)
DT
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4463.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy.
KW
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-0134001.
XX
PR 14-AUG-1997; 97US-005779P.
PR 08-NOV-1997; 97US-064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
PI Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
DR N-PSDB; ABN92163.
XX
PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
XX
PS Disclosure; SEQ ID 4463; 267pp; English.
XX
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX
SQ Sequence 2137 AA;

Query Match 7.4%; Score 237; DB 23; Length 2137;
Best Local Similarity 19.9%; Pred. No. 1.7e-05;
Matches 136; Conservative 134; Mismatches 316; Indels 96; Gaps 23;

QY 9 ESGPESVSSNOSMPTINGQJASNSE-----TKESTKASEA---SPSASSVSSMSFLS 60
DB 1079 ESDSTSTSLSESTSTSLGSGTASSTSDASSTSESDSTSESTSLSESLSTVSOSTSAS 1138
QY 61 SAKNALISLRDAIILKNKSSPTDLSQLEASTSTSTVTVRAADYDEAKSNFTAKSGLEN 120
DB 1139 TSESASTSTSES--ESNASTSLSGSLSTISDSTSTIS---DSASTS--TSESESDS 1190
QY 121 AKTLAEYETKADLMALQDMERLANSBPNNHTEEVNINKALEAQKDTIDKLKLVTL 180

PF 21-MAR-2001: 2001WO-US09180.
 XX
 PR 21-MAR-2000: 2000US-191078P.
 PR 23-MAY-2000: 2000US-206848P.
 PR 26-MAY-2000: 2000US-207727P.
 PR 23-OCT-2000: 2000US-242578P.
 PR 27-NOV-2000: 2000US-253625P.
 PR 22-DEC-2000: 2000US-257931P.
 PR 16-FEB-2001: 2001US-269308P.
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU;
 PI Yamamoto RT, Xu HH;
 PI WPI: 2001-611495/70.
 DR N-PSDB: AAS52197.
 XX
 PT New polynucleotides for the identification and development of
 XX antibiotics, comprise sequences of antisense nucleic acids -
 PS Example 3; Seq ID No 5834; 51pp; English.
 XX
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX Sequence 2437 AA:
 SQ
 Query Match 7.4%; Score 235.5; DB 22; Length 2437;
 Best Local Similarity 20.0%; Pred. No. 2.5e-05;
 Matches 141; Conservative 123; Mismatches 205; Indels 235; Gaps 29;
 OY 10 SGPESVSSNOSSMNPILNQIASNSETKASEASPSASSSVSSMFLSSAKNALISL 69
 DB 1413 NGDONIANNAKDKANAFVNSLNGLNQOQ-----DLAHKAINNADYSDVTIDVNNQID 1466
 OY 70 RDAILKNSSPPTSLSQL--EASTSTSTVTRYAAKDYDEAKSNFTAKSGLENAKTLAE 126
 DB 1467 NDAM-----ETLKHLDVNEIPNAEQTVNYQNAD--DNATKNFEDAK----- 1505
 OY 127 YETKAADLMAALODMERLANSDPSSNNHTEEVNNIKALEAQDITDKL-----KIV 178
 DB 1506 -----RLNLTNLNSDNT--NVNDINGAIOAVNDIAIHNLGCDRLQADAKK 1548
 OY 179 TLONOKSLTEVLKTTTDSA-----DOIPIAINSOLE-----INNSADQIIKDLERO- 224
 DB 1549 AIOSINQALANKLKEIASNATDQDLIAKNKAEEILANSIINNINATNSQAESOVOTAG 1608
 OY 225 NISYEAVLTNAGEVTKASSEAGIKLQALQSTVDAGDOSAAVLAQAQNNSPDNIATKE 284
 DB 1609 NHAIEGV--HANEIPAKIKIDANKDVQYQALIDEIDR-----NPLTIDK 1651
 OY 285 LIDAAETKYNF-LKQHTGTIDSPF--VKKAEQISQAKOKDIOETKPSGSDIPYGPSC 340
 DB 1652 EKQALDRINQILOGHGINNAMTEIEIOAKQAQALQADIDIKDLVKAKEADK----- 1705

OY 341 SASASGAAGALKSSNNGSRISLLDDVDNEMAATALOGFSRMIEQFVNNPATAKELOA 400
 DB 1706 -----ODVKKOYA-----LIDEIDQNPULTKEKOA 1732
 OY 401 MEAETIAMSQDLVGADGE-----LPAETIOAIKDALQALQOPSADGIATAMGOVAFNA 453
 DB 1733 LMDRIHQIILQO-----GHNDINNALLKKEIEIOAKQAQALQ----- 1769
 OY 454 AKVGGGSACTAGTVQNNVKOLYKTAFSSYSSSSYAAALSDGYSAKYKTLMSLYSESRGVQ 513
 DB 1770 -----DIDLK-----AKEDARNAIK--ALANAKRDQIN 1797
 OY 514 SAISQTPANPALSRSYRSRSGIESGSRASDASORAAETIVDSOTIGDVSRLQV-IDSLMS 572
 DB 1798 S-----NPDLTPEQAKAL-----KEIDEAKRALQNVENAOITDQNLNGLNLGLDIDRN 1847
 OY 573 TTV---SNPQANQ-----EEM-----QKLTASISKAPOFG-----YP 602
 DB 1848 THWVEVDEQPAVNEIFEARPEQILVNGELIVHRDQITFEQDILAHNLIIDQLSAEVIDTP 1907
 OY 603 AVONASDSIQKFAQLEREFDGERSLAESQENAFRKQAPFIQO 646
 DB 1908 STATISDSL--TAKVEVTILDGSKVIVVPYKVEKEKLSVVKQ 1948
 RESULT 29
 ID AAW98149
 AA AAW98149 standard; Protein; 2285 AA.
 AC AAW98149;
 XX
 DE 05-JUL-1999 (first entry)
 XX
 DE Bacillus subtilis metalloprotease YOMI.
 XX
 KW YOMI: metalloprotease; protease; textile; animal feed; detergent;
 KW Gram-positive bacterium.
 OS Bacillus subtilis.
 PN WO9914342-A1.
 PD 25-MAR-1999.
 PF 08-SEP-1998: 98WO-US18828.
 PR 15-SEP-1997: 97GB-0019636.
 PA (GENEV) GENENCOR INT BV.
 PA (GENEV) GENENCOR INT INC.
 PI Estell DA;
 PI WPI: 1999-229541/19.
 PI N-PSDB: AAX24980.
 PT Compositions containing metalloprotease from Gram-positive
 PS microorganism
 XX
 PS Claim 4; Fig 1A-O; 59pp; English.
 XX
 CC The present sequence is metalloprotease YOMI of *Bacillus subtilis*.
 CC The invention relates to the discovery of this previously unknown
 CC metalloprotease, uses of the metalloprotease in industrial and
 CC agricultural applications, and advantageous strain improvements
 CC based on genetically engineering a Gram-positive microorganism to
 CC delete, underexpress or overexpress the enzyme. Due to overall
 CC relatedness of the enzyme with *Pseudomonas lasa* protein, YOMI
 CC appears to be a member of the M23 metalloprotease family. The
 CC metalloprotease can be used in claimed cleaning compositions. The
 CC animal feed and compositions for the treatment of textiles. It may
 CC also be used for peptide hydrolysis, waste treatment and for
 CC cleaving recombinant fusion proteins. Expression vectors


```
Db 302 NNLISQITSATRVSGVDVTKQNAQHLDQAMASLQNGINNESQVKSSEKYROADTKQOEY 361
Qy 158 NNIKALEAOKDTIDKLKLVTLQONKSLTEVLKTTSDAOIPAINSOLEINKKSADQI 217
Db 362 DN---AITAAKAILNKSSTGPTAQN---AVEAALQVRNNAKDALNGDAKLIAQNAKOH 415
Qy 218 IKDLERQNISEAVLTNNGEVTKASSEAGIK-----LQO-----AL 253
Db 416 LGTLHTTTAQRNDLTN--QISQATNLGAVESVKONANSLDGMGNLQTAINDKSGTILAS 473
Qy 254 QSIYDAGQSOAAVLQ-----QONNSPDNIATKELIDAEETKYNELKOEHTGLTD 305
Db 474 QNFLDADQKRNAYNQAVSAETILNKOTGPN---TAKTAVBQALNNVNNNAKHALNGQN 530
Qy 306 SPLVYKA-----EEQISOAKDIOETKPSGS-----DPIYGPSSGSAASGSAAGA 351
Db 531 LNNAKQAALITANGASDINOKOKDALKAQANGAQRVSNAQOV-----OHNAETELNTAMGT 585
Qy 352 LKSSNNSGRISLLDDVDNEMAALQGFBSMIEQFNVNNPATAKEL-----QAMEAO 404
Db 586 LKHA-----IADKNTNTLAASSKYVNADSTKQNAVTTKYTNMAEHITSGPTTVYTTTSE 636
Qy 405 LTAMSDQLVGADGELPAE--IOAIK-----DALAQALKOPSDAGLATAMQOVAFAAAK 455
Db 637 VTAAANQVNSAKQELNGDEBLREAKONANTALDALTO-LNTPQAKKLEQVGO-----688
Qy 456 VGGSGAGTAGYQAM-----VKOLYTFARST---SSSSTAALSLDGYSAKYKTLNLSLY 505
Db 689 --ANLEEDQYVQTNQALNNNAKGLRDSIANETTVKTSQNTDASPPNQSTY---NSAV 743
Qy 506 SESRSGVQSAISQTPANPALSRVSRSIGESQGRSADASORAETIVRSQTLGDVYSRLQ 565
Db 744 SNAK-----GIINQTNPTMDTSATQ-----TTOYNNNAKNGLNGENLRNNAQNTAKQMLNLS 798
Qy 566 VLDSIMSTVSNPQANQEIIMOKLTASISKAPQFGYPAVQNSADSLQKFAQLER-----620
Db 799 HLTNNOKSAISS-QIDRAGHVSEVTAT-----KNATETELNTQMGLNEQAIHQ 845
Qy 621 -----EFVGERSLAESQENAFKOPAFIOQ 646
Db 846 NTWKOSVAKFTDADKAKKRDAYTNMAVSRAEDAILNK 878

RESULT 31
AAU37490
ID AAU37490 standard; Protein: 1048 AA.
XX
AC AAU37490;
XX
Df 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #1660.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
PN MO200170955-A2.
PD 27-SEP-2001.
PE 21-MAR-2001; 2001WO-US09180.
PF
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
```

```
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55349.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS
XX
PS Example 3; Seq ID No 13083; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 1048 AA:
```

Query Match 7.3%; Score 234.5; DB 22; Length 1048;
Best Local Similarity 20.5%; Pred. No. 9.3e-06;
Matches 154; Conservative 132; Mismatches 288; Indels 179; Gaps 32;

```
Qy 10 SGPEVSSNOSMNPITNGQIASNETKESTRKASASVSASSVSNFSLSAKALLSL 69
Db 201 AGVTFQVQSNATLDDQAMN-TLRQSIANKDARKASEDYDVAAND-----KQTAAYNNAAVA 253
Qy 70 RDAILNKNSSP-----TDSLQLEASTS-----TSTVTRVAAKDY-----DEAK 108
Db 254 AETIINANSNPEBNDSTTQAKAEQVNSKKTALNGDENITAAKQAKKTYLNTLTSTTDQK 313
Qy 109 SNF-----DTAKSGLENAKTLAEY-ETRMADLMAALQDMERLANS-----DPSNNHTEEV 157
Db 314 NNLISQITSATRVSGVDVTKQNAQHLDQAMASLQNGINNESQVKSSEKYROADTKQOEY 373
Qy 158 NNIKALEAOKDTIDKLKLVTLQONKSLTEVLKTTSDAOIPAINSOLEINKKSADQI 217
Db 374 DN---AITAAKAILNKSSTGPTAQN---AVEAALQVRNNAKDALNGDAKLIAQNAKOH 427
Qy 218 IKDLERQNISEAVLTNNGEVTKASSEAGIK-----LQO-----AL 253
Db 428 LGTLHTTTAQRNDLTN--QISQATNLGAVESVKONANSLDGMGNLQTAINDKSGTILAS 485
Qy 254 QSIYDAGQSOAAVLQ-----QONNSPDNIATKELIDAEETKYNELKOEHTGLTD 305
Db 486 QNFLDADQKRNAYNQAVSAETILNKOTGPN---TAKTAVBQALNNVNNNAKHALNGQN 542
Qy 306 SPLVYKA-----EEQISOAKDIOETKPSGS-----DPIYGPSSGSAASGSAAGA 351
Db 543 LNNAKQAALITANGASDINOKOKDALKAQANGAQRVSNAQOV-----OHNAETELNTAMGT 597
Qy 352 LKSSNNSGRISLLDDVDNEMAALQGFBSMIEQFNVNNPATAKEL-----QAMEAO 404
Db 598 LKHA-----IADKNTNTLAASSKYVNADSTKQNAVTTKYTNMAEHITSGPTTVYTTTSE 648
Qy 405 LTAMSDQLVGADGELPAE--IOAIK-----DALAQALKOPSDAGLATAMQOVAFAAAK 455
```

Db	649	VTAANAQVNSAKQELNGEDERLEREKQVANAANAIDALITQ-LNTPQKAKIKEVQQ-----	700
Qy	456	VGGSGAGACTGVQNM-----VKQLYKTAFFSST-SSSSVAALSDGYSAYKTLNSLY	505
Db	701	--ANRLEDEVQVQNGCALNNAMKGLDSDIANETTVKTSQNTYDASPNQSTY-NSAV	755
Qy	506	SESSSGVQSAISIQTPANALSRVSRSRISIESQGSADASQRAETIVRDSOTLGDVYSRLQ	565
Db	756	SNAR---GIINQNNPNTMTDSAITQO-TQVNNAKKGLGANGELRNQAONTAKQNTLIS	810
Qy	566	VLDLSMTSIVSNPANOEEIMQKLTASISKAPQGYPAVONSADSLQKFAALER-----	620
Db	811	HLTNQKSAITS-QIDRAGHYSEVTAI-----KNAATELNTQNGNLEQAIHQ	857
Qy	621	-----EFYDERSLAESQENAFRRKOPAFIQ	646
Db	858	NTVKQSVKFTDADKAKRDAYTNMVSRAEALLNK	890
RESULT 32			
AAU37464	AAU37464	standard; Protein: 560 AA.	
AC	AAU37464:		
DT	14-FEB-2002	(first entry)	
De	Staphylococcus aureus	cellular proliferation protein #1634.	
XX	Antisense: prokaryotic cellular proliferation protein;		
KW	antibiotic; antibacterial; drug design.		
XX	Staphylococcus aureus.		
OS	W0200170955-A2.		
PN	27-SEP-2001.		
PD			
XX	21-MAR-2001; 2001WO-US09180.		
PF			
XX	21-MAR-2000; 2000US-191078P.		
PR	23-MAY-2000; 2000US-206848P.		
PR	26-MAY-2000; 2000US-207727P.		
PR	23-OCT-2000; 2000US-242578P.		
PR	27-NOV-2000; 2000US-253625P.		
PR	22-DEC-2000; 2000US-257931P.		
PR	16-FEB-2001; 2001US-269308P.		
PA	(ELIT-) ELITRA PHARM INC.		
PI	Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;		
PI	Yamamoto RT, Xu HH;		
XX			
DR	WPI: 2001-611495/70.		
XX	N-PSDB; AAS55323.		
PT	New polynucleotides for the identification and development of		
PT	antibiotics, comprise sequences of antisense nucleic acids -		
XX			
PS	Example 3; Seq ID No 13057; 51pp; English.		
CC	The invention relates to antisense inhibitors of genes essential to		
CC	prokaryotic cellular proliferation, their use in identifying the		
CC	genes, their use in the discovery of novel antibiotics, the essential		
CC	genes themselves and the encoded proteins. The prokaryotes used are		
CC	Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella		
CC	pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The		
CC	invention is also useful for the identification of potential new targets		
CC	for antibiotic development. The antisense nucleic acids can also be used		
CC	to identify proteins used in proliferation, to express these proteins,		
CC	and to obtain antipodes capable of binding to the expressed proteins.		
CC	The proteins can be used to screen compounds in rational drug discovery		
CC	programmes. The antisense nucleic acid sequence is also useful to screen		

[illegible]

```

PN FR2807446-A1.
XX
XX 12-OCT-2001.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX 11-APR-2000; 2000FR-0004630.
XX
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.
XX
XX Bolotine A, Sorokline A, Renault P, Ehrlich SD;
XX
XX WPI; 2002-043418/06.
XX
XX New nucleotide sequence useful in the identification of Lactococcus
XX lactic and related species -
XX
XX Claim 6; SEQ ID NO 870; 2504bp; French.
XX
XX The present invention is related to a Lactococcus lactis nucleotide
XX sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
XX nucleic acid sequence is useful in the detection and/or amplification of
XX nucleic acid sequence, particularly to identify Lactococcus lactis or
XX related species. The proteins of the invention are useful for the
XX biosynthesis or biodegradation of a composition of interest. The
XX invention helps research in lactic bacteria, particularly useful in the
XX production of yogurt and cheese.
XX
XX Note: The sequence data for this patent is based on equivalent patent
XX WO200177334 (published 18-OCT-2001) which is available in electronic
XX format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1063 AA:
XX
XX Query Match 7.3%, Score 234; DB 23; Length 1063;
XX Best Local Similarity 22.4%; Pred. No. 1e-05;
XX Matches 160; Conservative 108; Mismatches 285; Indels 160; Gaps 29;
XX
OY 13 ESVSSNOSSMNP-----INGQIANSSEKSTKASEAS--PSASSSVSMFL-- 59
DB 146 DSSSSNDGNSISLSSSNAADSVSAVGSOSTSSGVLSSESAIDSGIASVSGSSENL 205
OY 60 -----SSAKNALISLRDAILNKNSSPTDLSQ-----LEASTSTSTVTVAAKD-YDEAK 108
DB 206 VGNSSASASAAVASFMTILATNPSMVPMLTGALAAAPATSSGAILNTTIGDLVNGAI 265
OY 109 SNFDTAKGLENA-KTLAEYETK-MADLMAALQDMERLANSOPSNHTEEVNNIKALEA 166
DB 266 ST--VGISGLANIFSTLCTFNIPGHTAAALNGVEQIVNI-----VGNIOEA--- 311
OY 167 QKDTIDKLKLVTLONOKKSLTEVAKTTDSADQIPAINSOLEINKNSADQITKDERONI 226
DB 312 -----AANPAFLNTERKS-----AGLDVSGQIPLVGGQIAAFAAAT-----PSM 350
OY 227 SYEAVLT--NAGEVIAKASSEAGIKGLQALOSIVDAGDSQAVALAOONNSPDNI--AAT 282
DB 351 SPAAMLTFLNPTIPGLSSIRGASL--VLSPLVLSAISTVTSIVN-QLNTTISNALGVN 407
OY 283 KELIDAEFTKYNELKOEHTGLTDSPLVKKAEBEQISOAKQIDQIEIKPSGSDIPVGPSC-- 340
DB 408 FDLDTLVLOGNDLVNVLAVGAVNSAIRNV-QQIAMSQ-----LSPTISNIPLVCTGVN 461
OY 341 -----SAAASGSAAGALKSSNNSGRISILLDDVDNEMAAILQGFERSMIEFPNV-- 389
DB 462 VLSPTLNLNLTGASGEVA-----NLGVSSLLDQVNNNSLGNLISLSTALATIENTLQ 514
OY 390 -----NNPATAKE-----IQAMEAQLTAMSDQLVGADGELPAEQIAIKDALQALQK- 436
DB 515 NSLNSFGNLPAGASDILNQLONALNINIVESATGIYNNRPG-LGAIENLSMTISQI 573
OY 437 -----PSADGLATAMGOVA-----FAAAKVGGSAGTAGTYVOM 469
DB 574 PNNINNVNNAIINGITTTIINSLTSPVGASTVNPNSANSOSSSSASSSSASSSSSTSS 633

```

```

OY 470 NVKOLYKT-----AFSSSTSSSYAALSDGYSAKYKTLNLSYSESRSGVOAISOTANPA 523
DB 634 NVSSNTSSNSBRANTSSSTSNASSSSSSSESSA-----SSSNSSESSVASSSSVDSSQS 688
OY 524 LRSYSRSGIESQGRSADAPRAETIYRDSOTLGDVYSRLQVLDLSMTIVSNPOANOE 583
DB 689 SSAGVNSSSSSSAEGSSASSNSSESSVASSSS-----VDSQS---SSAGVNS 734
OY 584 EIMOKLTASISKAPQFGYPAVONSADSLQKFAQLEREFVDEGERSLASQENA 636
DB 735 SSSSESSASSNSSESSGVASSSVDSOSSAGVNSSSSSSAEGSSASSNS 787

RESULT 34
ID AU34207
ID AU34207 standard; Protein: 2025 AA.
XX
XX AU34207;
XX
XX 14-FEB-2002 (first entry)
XX
XX Staphylococcus aureus cellular proliferation protein #483.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX
XX 23-MAY-2000; 2000US-206848P.
XX
XX 26-MAY-2000; 2000US-207727P.
XX
XX 23-OCT-2000; 2000US-242578P.
XX
XX 27-NOV-2000; 2000US-253625P.
XX
XX 22-DEC-2000; 2000US-257931P.
XX
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KI, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX
XX N-PSDB; AAS52066.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5703; 511bp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at

```

```

CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2025 AA:
XX
Query Match 7.3%; Score 233; DB 22; Length 2025;
Best Local Similarity 19.8%; Pred. No. 2.7e-05;
Matches 141; Conservative 129; Mismatches 275; Indels 166; Gaps 25;
QY 10 SGPESVSSNOSMPTIINGIASNETKSTKSEASPSASSSVSSWSPFLSAKNNLISL 69
DB 1078 NGQONLANKKDKNAFVNSLNGLNQOQ-----DLAHNAINNADIVSDTDLVNNQIDL 1131
QY 70 RDAIILKNSPDSLSOL---EASTSTSTVTRVAANDYDPAKSNFTPASGLENAFTLAE 126
DB 1132 NDAM-----ETLKHLYDNEIPNAEQTVYQNAD--DNAKTNFDDAK----- 1170
QY 127 YEKRMADLMAALQDMERLANSNHTTEVNNIKKALEOKRTIDKL-----KLV 178
DB 1171 -----RLANTLLNSDNT--NVNDINGAIGQVDAIGNLNGEORLQEAQDK 1213
QY 179 TLQONKSLTEVLKTTDSA-----DOIIPAINSOLEINKNSADQIIDLERQNISYEAVLT 233
DB 1214 AIONVAKVLADKLEIASNATDQDKLIAKNAEELANSTINNINKATSQDVSYQTAG 1273
QY 234 N-----AGEVIRKASSEAGIKIGQALQSIYDAGDOSQAAYVLOAQOONSPPDNIATKELI 286
DB 1274 NQALIEGVHANEIIPKAKIDANKVDKQVQALD-----DEIGRPNPLT 1314
QY 287 D-----AETKRVNE-LKQERTGLTDS---PLVKAEEQISOAKDIOETKSSGSDIPIVCP 338
DB 1315 DKRKQALKRINOIILQOGHNDINNALTKEAIEQAKERLQALODIDLVYAKED----- 1368
QY 339 SGSAAGASAGALKSSNNSGRISL-----LLDQVDNEMAAIATLQGFSS--M 383
DB 1369 -----AKNKIKALANAKRKQIINSNDLPTEQAKALKKEID--EAEKALQONVERAQT 1418
QY 384 IEQFN-----VNN-----PATAKELQAMEAQLTANSDQLVGADG-----ELP 420
DB 1419 IDQLNGLNLGLDDIRNTHWEVDQPAVEISEATPEQLLVNGELIVHRDITTEQDVL 1478
QY 421 AELQATKDALAQLKQPS-----ADGLATAMGOVAFPAAKYGGSSAGTAGVQANVQLYK 476
DB 1479 AHNINLIDQLTAEVYIDPPTATTISDSTAKAEVTLIDGSKV-----IVNVPAVVERK 1529
QY 477 --TAFSTSSSYAALSDGYSAYKTLNLSYSESRSQVSAISQFANPALRSVSRSQIE 534
DB 1530 ELSVAVQQAIESEIENAVQOKINEINNSVLTLEQKEALAEVKKLQQAIDHNNAPDVH 1589
QY 535 SOGRSADASORAAETTVRDSQTLGDVYSR--LQVLDLSLMTSTIVSNPOANOEIMOKITLST 593
DB 1590 SVEEIQOQEQAHIEQFPNEQFTTEQAKSNAIKSIDAIIQHMIDEIKARTDULTQEKQEAL 1649
QY 594 SKRPGGYPAVQ--NSADSLQKFAQLERFVNGERSLAESQMAARKQQA 642
DB 1650 AKLNLQKEQAIOIQRASIDETTELQEQAKQAMKANPTAKELARRKQEA 1700
RESULT 35
AA034143
ID AA034143 standard; Protein; 2086 AA.
XX
XX AA034143;
XX
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #419.
XX
XX Antisense; prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX

```

```

XX
XX 27-SEP-2001.
XX
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX
XX 23-MAY-2000; 2000US-206848P.
XX
XX 26-MAY-2000; 2000US-207727P.
XX
XX 23-OCT-2000; 2000US-242578P.
XX
XX 27-NOV-2000; 2000US-253625P.
XX
XX 22-DEC-2000; 2000US-257931P.
XX
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELITR-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GT;
PI Yamamoto RT, Xu HH;
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52002.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5639; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2086 AA:
XX
Query Match 7.1%; Score 228; DB 22; Length 2086;
Best Local Similarity 22.0%; Pred. No. 5.7e-05;
Matches 166; Conservative 106; Mismatches 259; Indels 222; Gaps 35;
QY 25 IINGQIANSN-----TKES-----TKASEASPSASSVSSNLSAKNALI 67
DB 1221 ILNKQGSNSDKAAVDRAILOQVSTKDALNGDAKLAEEAQAQONIGTINHNTMAORTAL 1280
QY 68 -----SLRDAI--LNKNSPTD--SLSOLEASTS--TSTVTRVAAKDVEAKSN--F 111
DB 1281 EGOINQATVYDGVNTYKTNANTLIDGAMNSILOGSINRKDATLRNQNTLDADESRNAYTQA 1340
QY 112 DPAKSGLENAKTLAEYETKADL-----MAALQDMERLANSNHTTEVNNIK 161
DB 1341 VTAAGELNKQGTG--NTSKADVDNALNTYTRAKAALNGAENLRNRTKSATNT--INGLP 1396
QY 162 KALEAKDKTI-----DKLNLKLVT-----LQONKSLTEVLKTTMS 196
DB 1397 NLTOLOKMDLKQVEQAQVAVAGVNGYKDGKNTLMTALGALRTSIQDNTTKT-----S 1449
QY 197 AQDIIPAINSOLEINKNSAQIIDLERQNISYAVLTVNGEVIKASSEAGIKGQALQSI 256
DB 1450 QNTLDASDS--NKN-----NYTAVNANNGVITVNNPNMD--ANALNG 1490
QY 257 VDAQDSQAAYVLOAQO-----NNSPDNIA-----ATKELIDAAET----- 291

```

```
Db 1491 ANOVTTKAALNGAQNLAQAKTTAKNNLNLTISINNAOKDALTRSIDGATTVAGVNOETA 1550
Oy 292 KVNELKOE-----HGLTDSPLYKKAEEIOISQAOKDIOEIKPSGSDIPYVPSGSA--SA 345
Db 1551 KATELNNAHMSLQNGIND-----ETQTKQTKYLD-----AGPSKSAVQQA 1592
Oy 346 GSAAGAL--KSSN-----NSGTSILLDDVDNEMAAIALOGFRSMIEGFN 388
Db 1593 VNAKAILITKASGONVDAKAAVEQALQNVNSTKTALNGAKLLEAKAAKQTLGTLTHINN 1652
Oy 389 VNNPATAKEL-QAMEAQ-----LTAMSDQLVGADGELPAEIOAIKDALAQAOKPSADGIA 443
Db 1653 AORTALDNEITQATNVEGVNIVKAKAQOLDGAMGQLERSIRHKDPTTLOSQNVQDADDAKR 1712
Oy 444 TAMGOVAFPAAKVGGSGAGTAGTVOMNKKLYTAFSSTSSSYAAALSDGYSAYKTINS 503
Db 1713 TAYSOAVNAATAI-----LNKTAGGNTPRADVERAQAQVTAQNTALNG 1755
Oy 504 LYSERSGVSQSAISOTANPALIS-----RSVSRSGIESGREGSADASQRAEITIVDS 554
Db 1756 IONLERA-----KOANTFALTINASDLNFKOEALKAQVTSAGRVs-----AANGVEHTA 1804
Oy 555 QTLGDVYSRLQ-----VLDLSMTSTIVSNPOANOEIIM-OKLTA-----SISKAPQFGYPA 603
Db 1805 TELNTPMTALMHAIDAKETKASGVNVNADANKRQAYDEKVTAAENIYSGPTPTLTTPSD 1864
Oy 604 VONSADSLQKFRAPQDEREVDGERSLAESQENA 636
Db 1865 VTNATQVNTNAKTOL-----NGNHNLEVAKQNA 1892
```

RESULT 36

```
AAU37017 standard; Protein; 5795 AA.
ID AAU37017
XX AAU37017;
XX
AC
XX
XX 14-FEB-2002 (first entry)
Dt
XX
XX Staphylococcus aureus cellular proliferation protein #1187.
De
XX
XX Antisense: prokaryotic cellular proliferation protein;
Kw antibiotic; antibacterial; drug design.
Kv
XX
OS Staphylococcus aureus.
XX
XX MO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KJ, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
XX
XX N-PSDB: AAS54876.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12610; 511pp; English.
```

```
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from Wipo at
CC ftp.wipo.int/pub/published_pcl_sequences.
XX
XX Sequence 5795 AA:
```

Query Match 7.1%; Score 228; DB 22; Length 5795;

Best Local Similarity 22.0%; Pred. No. 0.00022;

Matches 166; Conservative 106; Mismatches 259; Indels 222; Gaps 35;

```
Oy 25 IINQOIASNSE-----TKES-----TKASESPASASSVSWSLSSAKNALI 67
Db 3498 ILNKQTSNSKRAAVDRALQGVTSYKDALNGDAKLAEEAKAAKQNLGTLNHTNAQRTAL 3557
Oy 68 -----SLRPAI-----LNKSSPTD-SLSQLEASTS--TSTYTRAAKQYDEAKSN----F 111
Db 3558 EGOINQATTVGCVNTKTNATLTDGAMNSLOGSINDKDTLNMONYLDADSKRAAYQA 3617
Oy 112 DTKASGLENAKTLAEYETKMADL-----MAALODMERLANSDPSNNHTEEVNINIK 161
Db 3618 VTAAGLILNKQTCG--NTSKADVDAALNTVTRAKAALGAELRLTKTSATNT--INLIP 3673
Oy 162 KALEAQKDTI-----DKLNKLYT-----LONOKSLTEVLKTTDS 196
Db 3674 NLTOLOKDNLKHQVBOAQNVAQVNGVCKDKGNTLNTAMGALRPTSIQNDVTKT-----S 3726
Oy 197 ADQIPAINSOLEINKNSADQIIKDERONISEVNLVTNAGEYIKASSBAGIKLQALQSI 256
Db 3727 QNYLDASDS---NKN-----NNTAVNANGVINTVNPND-ANAINGM 3767
Oy 257 VDAGDOSQAAYLQAOQ-----NNSPDNIA-----ATKELIDAET----- 291
Db 3768 ANOVTTKAALNGAQNLAQAKTTAKNNLNLTISINNAOKDALTRSIDGATTVAGVNOETA 3827
Oy 292 KVNELKOE-----HGLTDSPLYKKAEEIOISQAOKDIOEIKPSGSDIPYVPSGSA--SA 345
Db 3828 KATELNNAHMSLQNGIND-----ETQTKQTKYLD-----AGPSKSAVQQA 3869
Oy 346 GSAAGAL--KSSN-----NSGTSILLDDVDNEMAAIALOGFRSMIEGFN 388
Db 3870 VNAKAILITKASGONVDAKAAVEQALQNVNSTKTALNGAKLLEAKAAKQTLGTLTHINN 3929
Oy 389 VNNPATAKEL-QAMEAQ-----LTAMSDQLVGADGELPAEIOAIKDALAQAOKPSADGIA 443
Db 3930 AORTALDNEITQATNVEGVNIVKAKAQOLDGAMGQLERSIRHKDPTTLOSQNVQDADDAKR 3989
Oy 444 TAMGOVAFPAAKVGGSGAGTAGTVOMNKKLYTAFSSTSSSYAAALSDGYSAYKTINS 503
Db 3990 TAYSOAVNAATAI-----LNKTAGGNTPRADVERAQAQVTAQNTALNG 4032
Oy 504 LYSERSGVSQSAISOTANPALIS-----RSVSRSGIESGREGSADASQRAEITIVDS 554
Db 4033 IONLERA-----KOANTFALTINASDLNFKOEALKAQVTSAGRVs-----AANGVEHTA 4081
Oy 555 QTLGDVYSRLQ-----VLDLSMTSTIVSNPOANOEIIM-OKLTA-----SISKAPQFGYPA 603
```



```

PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX DR N-PDB; AAS52179.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 5816; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2478 AA:
XX
XX Query Match 6.9%; Score 222; DB 22; Length 2478;
XX Best Local Similarity 19.6%; Pred. No. 0.00016;
XX Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;
XX
XX QY 14 SVSSNOSNMPIINGQJASNSETKESRKASPSASSVSWSFSLSAKALLSLDAI 73
XX DB 535 AVENKMDALQOOVNSQVDSNHYT-----TASIAEY-----NKLKQQADTI 574
XX QY 74 LNKNSPTDLSQLEASTSTSTVTVAAKDYDEAKSNFDTAKSGLENAKTLAEYETKAD 133
XX DB 575 LN-----EDANHVKTANRASQADIDGLVTKLQAL--IDNQAIAEIDTKAQE 620
XX QY 134 LMAALQDMERLANSDP-----SNHTEVYNNIKKALEAKQDTIDKLKLVTLQONKS 186
XX DB 621 KVTAAQOSKRYTQDEVALVTKINNDKNNAIAEINKQTTAQGVTEKONGIAVLEQD-V 678
XX QY 187 LTEVLTQTDASADQIPAI-----NSOLEIKKNSAD-----QIYKDLF--KQNT 226
XX DB 679 ITPTVKQAKODIIQAVTTTKRQKQIKSNASIQDEKDVANDKIGKIEKRAIDAAATMA 738
XX QY 227 SYEAVLTNAGEVIRKASSEAGIKLQALQSIYDAGDQSOAVALQAOONNSPNIATKELI 286
XX DB 739 QVEAIKTKAINDIQTPATPAKAAALEFDE-----VYQAIQDQALNPPTTNEEV 790
XX QY 287 DAAETKYNELKQEHGTGLTSDPLVKABEQISOAKDIOEIKPSGSDIPIVPGSGSAAAG 346
XX DB 791 AEAERINAAKV--SGV-----KAIEATTTAQ--DLEPYK-----821
XX QY 347 SAAGALKSSNNSGRISILLQDVNDENMAIALQGRSMIEQFN--VNNPATAKEIQAMEQQL 405
XX DB 822 -----NEBISKIENITTSOTQTKMDA-----YNEKQAATATKAKQANATVS- 860
XX QY 406 TAMSQDLVGADGEL-PAEIOAIKDALQALQPSADGLATAMQGVAPFAAAKVGGSAGTA 464

```

```

DB 861 NATNEEVAEADAADAAQKQGLHDIQVYKSKQEVADTKSKYLDKINAIQIQ-----AKVK 915
XX
XX QY 465 GTYQAMNKQLYKTAFFSTSSSYAAALSDGYSAKTTINSLXSEKSGVQ-----513
XX DB 916 PAADTEVENAYNTRKQEIQNSN-ASTTEKQAAATEIDTKQOEARTIMDANNTSDVTTA 974
XX QY 514 -----SAISQIOT-----ANPALSRSYSRSGICSGRSADASQRAAEITVRSQTL 557
XX DB 975 KDNSTAIAMINOYQAAATTTKSDAKAEIQAQKASERKATIAEMDSTTEEQAA-----KQVYDQ 1030
XX QY 558 GDVYSRLQVLDLSMSTIVSNPQANQEIIMQKLTASISKAPQFGYPAYVANSLSQKFPAAQ 617
XX DB 1031 AVYTAANMIDNMAANNVDNAKNTNEATIAITPDPANVKP-----AAQALADKV-----Q 1081
XX QY 618 LERFVUGENSLAESQENARFRKOPAFIQOVLVNTA 652
XX DB 1082 AOETAIQDNGNG-SYTEKKAQKQOVQTEKTTADAA 1115
XX
XX RESULT 39
XX AAU37374
XX ID AAU37374 standard; Protein; 2478 AA.
XX
XX AAU37374;
XX
XX 14-FEB-2002 (first entry)
XX
XX Staphylococcus aureus cellular proliferation protein #1544.
XX
XX Antisense: prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Staphylococcus aureus.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI; 2001-611495/70.
XX DR N-PDB; AAS5233.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 12967; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:22:42 ; Search time 72 Seconds
(without alignments)
1888.765 Million cell updates/sec

Title: US-10-007-693-139

Perfect score: 3204

Sequence: 1 MHHHHHMSPESSVSNQS.....PAFIQVLVNIAISLESGYLS 660

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2943	91.9	647	16	084627	084627 chlamydia t
2	894	27.9	651	16	0927H7	0927H7 chlamydia p
3	453	14.1	715	2	046166	046166 chlamydia p
4	272.5	8.5	3381	2	09KX33	09KX33 streptococ
5	272	8.5	2283	2	08VO99	08VO99 streptococ
6	268	8.4	2271	16	0990Y4	0990Y4 streptococ
7	261	8.1	6713	16	0990Y4	0990Y4 streptococ
8	261	8.1	6713	16	0931R6	0931R6 streptococ
9	237.5	7.4	2016	5	09B1T0	09B1T0 plecteurys
10	235	7.3	2285	16	064046	064046 bacterioph
11	235	7.3	1063	16	031976	031976 bacillus su
12	234	7.3	1063	16	09CH86	09CH86 lactococcus
13	233	7.3	1086	16	09A0H0	09A0H0 streptococ
14	233	7.3	1344	2	049545	049545 mycoplasma
15	231	7.2	1365	2	049525	049525 mycoplasma
16	230	7.2	2178	2	09KWR3	09KWR3 streptococ

17	228	7.1	1086	2	0938K3	0938K3 streptococ
18	228	7.1	3072	2	0939N5	0939N5 streptococ
19	227	7.1	2478	2	091CH2	091CH2 streptococ
20	226.5	7.1	969	5	09ND19	09ND19 plasmodium
21	226	7.1	1795	2	091CJ9	091CJ9 streptococ
22	225.5	7.0	1302	2	049547	049547 mycoplasma
23	225	7.0	1404	2	08RJN9	08RJN9 mycoplasma
24	224.5	7.0	1790	3	007380	007380 saccharomy
25	222	6.9	1327	2	09X7M2	09X7M2 streptococ
26	222	6.9	2478	2	09RL69	09RL69 streptococ
27	221.5	6.9	924	5	015738	015738 dictyostell
28	220.5	6.9	1627	5	096200	096200 giardia lam
29	219.5	6.9	1833	5	09VM67	09VM67 drosophila
30	219.5	6.9	2139	5	007569	007569 entamoeba h
31	219	6.8	1514	5	085Y55	085Y55 drosophila
32	218.5	6.8	2570	2	092FE9	092FE9 streptococ
33	218	6.8	4776	16	097P71	097P71 streptococ
34	217	6.8	1935	5	044934	044934 loligo peal
35	217	6.8	2297	3	09HGK6	09HGK6 candida alb
36	216.5	6.8	2055	2	085472	085472 abiotrophia
37	215.5	6.7	2124	16	098M03	098M03 rhizodium 1
38	214	6.7	716	2	049526	049526 mycoplasma
39	214	6.7	1072	16	09CF64	09CF64 lactococcus
40	213	6.6	1489	5	087805	087805 drosophila
41	213	6.6	1794	9	0971A7	0971A7 bacterioph
42	213	6.6	2186	16	099PB0	099PB0 streptococ
43	212	6.6	962	2	049546	049546 mycoplasma
44	212	6.6	1489	5	09W1R3	09W1R3 drosophila
45	212	6.6	1690	5	044929	044929 drosophila

ALIGNMENTS

RESULT 1

ID 084627 PRELIMINARY: PRT: 647 AA.
AC 084627:
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CHLPN 76 kDa homolog.
GN CT622.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UM-3/CX:
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammell C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RT Science 282:754-759(1998).
DR EMBL: AE001333; AAC68226.1; -.
DR PHCI-2DPAGE: 084627; -.
KW Complete proteome.
SQ SEQUENCE 647 AA; 68525 MW; COD1AC2D74473625 CRC64;

Query Match 91.9%; Score 2943; DB 16; Length 647;
Best Local Similarity 93.9%; Pred. No. 7e-110;
Matches 613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;
QY 8 MESSGESVSSNNSNNPIINCOIANSSEKTSKRSSEASASSVSSWSPFLSSAKNALI 67
DB 1 MESSGESVSSNNSNNPIINCOIANSSEKTSKRSSEASASSVSSWSPFLSSAKNALI 60
QY 68 SLRDAILNKNSSPTDSLQLEASTSTSTSTVTVRAARDYDEAKSNFDTANSGLLENATVLA 127
DB 61 SLRDAILNKNSSPTDSLQLEASTSTSTSTVTVRAARDYDEAKSNFDTANSGLLENATVLA 120

QY 128 ETRMADLMAALODMERLANSDBSNHTEEVNNIKKALEAQKDTIDKLKLVTLQNONKSL 187
 DB 121 ETRMADLMAALODMERLA-----KOKAEVTRKEKLOEKOEVIDKLQVLEKLEKONQTL 174
 QY 188 TEVLKTTSDADQIPAINSOLEINKNSADQIIKDLERONISYEVLVFNNAEVIKASAEAGI 247
 DB 175 KETLVTTSDADQIPAINSOLEINKNSADQIIKDLERONISYEVLVFNNAEVIKASAEAGI 234
 QY 248 KLGQALQSLIVDAGDQSOAAVLQAQONNSPDNIATKELIDAAETKVNELKOEHTGTLTDP 307
 DB 235 KLGQALQSLIVDAGDQSOAAVLQAQONNSPDNIATKELIDAAETKVNELKOEHTGTLTDP 294
 QY 308 LYKKAEEQISOAKDIOEIKPSSGSDIPYVPGSSAASAGSAGALAKSSNNSGRISLLDD 367
 DB 295 LYKKAEEQISOAKDIOEIKPSSGSDIPYVPGSSAASAGSAGALAKSSNNSGRISLLDD 354
 QY 368 VDNEMALAIQGRSMIEQFNVNNPATKELQAMEAQLTAMSQVLGAGGELPAEIQAIK 427
 DB 355 VDNEMALAIQGRSMIEQFNVNNPATKELQAMEAQLTAMSQVLGAGGELPAEIQAIK 414
 QY 428 DALAQALKOPASDGLATANGOVAFAAAKVGSGAGTAGTVQNMVKOLYKTAFFSSTSSSY 487
 DB 415 DALAQALKOPASDGLATANGOVAFAAAKVGSGAGTAGTVQNMVKOLYKTAFFSSTSSSY 474
 QY 488 AALSDGYAYKTLNLSLYSESRSGVQSAISQTPANPALSSVSRSIGTESQGRSADASQRAA 547
 DB 475 AALSDGYAYKTLNLSLYSESRSGVQSAISQTPANPALSSVSRSIGTESQGRSADASQRAA 534
 QY 548 ERTVRSQTLGIVYSRLQVLDLSMTITVNPQANOEIMOKLTASISKAPQEGYPAVQNS 607
 DB 535 ERTVRSQTLGIVYSRLQVLDLSMTITVNPQANOEIMOKLTASISKAPQEGYPAVQNS 594
 QY 608 ADSLQKFAQLEREFEVGERSLAESQENAFKOPAFIQOVLVIAISLFGSYLS 660
 DB 595 ADSLQKFAQLEREFEVGERSLAESQENAFKOPAFIQOVLVIAISLFGSYLS 647

RESULT 2
 ID 0927H7 PRELIMINARY: PRT: 651 AA.

AC 0927H7
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE CHLPN 76 kDa HOMOLOG_1 (CT622).
 GN CPN0728 OR CPJ0728 OR CP00128.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CM1029;
 RC MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Jammel C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomics of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RC MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heideberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry C.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RC MEDLINE=20330349; PubMed=10871362;

RA Shitai M., Hirakawa H., Kimoto M., Tebuchi M., Kishi F., Ouchi K.,
 RA Shida T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CM1029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL: AE001654; AAD18867.1; -;
 DR EMBL: AE002165; AAF37914.1; -;
 DR EMBL: AP002547; BAA98935.1; -;
 DR PHC1-2DPAGE: 0927H7; -;
 DR TIGR: CP0018; -;
 KW Complete proteome.
 SQ SEQUENCE 651 AA; 68217 MW; 47AE6C3FF2FF0123 CRC64;

Query Match 27.9%; Score 894; DB 16; Length 651;
 Best Local Similarity 34.6%; Pred. No. 2,7e-28;
 Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

QY 11 GPESVSSNOSNMPTINGOIASNSEFTKESTKA-----SEASP--SASSSVSSKFLSSAK 63
 DB 6 GPEPIDETERTPADLSAOGLEASANKSAEORIGAEAKPRESKTDSVERVSILRSAY 65
 QY 64 NALISLRD--ALINKSSPTDSLQLEASTSTVTRVAAKDYDEKSNFPAKSGLENA 121
 DB 66 NALMSLADKLGIASSVSSSTSRs-ADVSTRTATPTPPPTFDYKTAQATAYDTFTS 124
 QY 122 KTLAEYETRMADLMAALODMERLANSDBSNHTEEVNNIKKALEAQKDTID-----K 173
 DB 125 TSLADIQALVSLQDAVTNINIKDTAATD-----EEI-----AIAAMEWKNDAYVQAG 173
 QY 174 LNKLVTLQNONKSLTEVLKTTSDADQIPAINSOLEINKNSADQIIKDLERONIS--SYTA 230
 DB 174 ITELAKYASDNQALIDLSLKLTSFLLQALLOSVANNNKKAELKEMODNPVPGKTPA 233
 QY 231 VLTNAGEVIRKASEAGIKTGQALQSLIVDAGDQSOAAVLQAQONNSPDNIATKELIDAAE 290
 DB 234 IAOGLVDQTDATATQTEKQGNALRDYFAGQNSGAVENAKSSNLSINSDAKAATATAK 293
 QY 291 TKVNELKOEHTGTLTDSPLVKAEEQISOAKDIOEIKPS-GSDIPIVGPS-GSASAGSA 348
 DB 294 TQIAEAOKK---FPDPSILQEAQOWIOAEKDKNKTKPADGSGVPRPGITVSGSKQGS 350
 QY 349 AGALKSSNNGRISLLDDVDNEMALAIQGRSMIEQFNVNNPATKELQAMEAQLTAM 408
 DB 351 TGS1-----RVSMLLDAENETASILSGFQRMTHMTEPNDSQAQOEIAQAARAA 403
 QY 409 SQGLVADGELPAEIQAIKDALAQAL-KOPASDGLATANGOVAFAAAKVGSGAGTAGV 467
 DB 404 K--AAGDSAAAALADAKALEALGKAGQOQGIINALGQIASAIVSAGVPPAAASST 460
 QY 468 QMNVKOLYKTAFFSSTSSSYAALSDGYAYKTLNLSLYSESRs-GVQSAISQTPANPALSR 526
 DB 461 GGSVVKOLYKT--SKTSGSDYKQISAGYDAYKSINAYGRARDATRDYINNVSTPALTR 518
 QY 527 SVSRGIESQGRSADASQRAETIVRDSQTLGDVYSRLQVLDLSMTITVNPQANOEIM 586
 DB 519 SVPRARTEARG-PEKTDQALARIYSGNSRTLGDVYSQVSAISQTPANPALSR 577
 QY 587 OKITASISKAPQEGYPAVQNSADSIQKFAQLEREFEVGERSLAESQENAFKOPAFIQO 646
 DB 578 OKITSAVTKPPQGYGYVQSLSDSTOKFTAKLESIFAESGRTAETKALSFETNSLFTIOO 637
 QY 647 VLVNIALSFGYL 659
 DB 638 VLVNIGSLVGYL 650

RESULT 3
 ID 046166 PRELIMINARY: PRT: 715 AA.

AC 046166
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE 76 kDa protein.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. *Chlamydia*.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94156481; PubMed=7509320;
RA Perez-Melgosa M., Kuo C.-C., Campbell L.:
RT "Isolation and characterization of a gene encoding a Chlamydia
pneumoniae 76-kilodalton protein containing a species-specific
epitope."
RL Infect. Immun. 62:880-886(1994).
DR EMBL; L23921; AAA23117.1;
SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 14.1%; Score 453; DB 2; Length 715;
Best Local Similarity 28.9%; Pred. No. 1.1e-10;
Matches 136; Conservative 94; Mismatches 193; Indels 48; Gaps 14;

OY 11 GPESVSNOSNMPIINGQIASNSETKETKA-----SEASP--SASSSVSSMSPLSAK 63
DB 262 GGPIDERTERTPPADLSAQGLEASANKSAEORTAGAKRKESKTSVERKSTLRSAV 321
OY 64 NALISLRD--ALLNKNSEPTDSLQLEASTSTVTVRAAKDYDEAKSNFTAKSGLENA 121
DB 322 NALMSLADKLGIASSNSSSTSR--ADVDTATATAPPTPEFDYKQTQAOTAYDTIFTS 380
OY 122 KTLAEYETKMDLMAALODMERLANSDESNHTEEVNNTKKALEAKOTIID-----K 173
DB 361 TSLADIGALVSLQDAVNNIKIPTAATD-----EET-----AIAEMETKMDAVKVAQ 429
OY 174 LNKLVTLQNMKSLTEVLTCTDSADQIPAINSOLEINKNSADQIITKLERONI---SYEA 230
DB 430 ITELKAYASDNQAIIDSLGKLTSPDLQALLQSVANNKKAELKEMDNVYREKTPA 489
OY 231 VLTNAGEVTKASSEAGIKLQALQSTVDAGDOSQAAVLQAOQNSPDNIATKELIDAE 290
DB 430 IQASIVDQDATATQIEKGNALIRDAVFAGQNASGAVENAKSNISINISAKAATATAK 549
OY 291 TVVNLKQHTGLTSPLYKKAEBQISOAKDIOETKPS--GSDIPYVGS--GSASAGSA 348
DB 550 TQIAEAKQK--FPDSPIQEAQVNTQAEKDKLNKPKGSDVPPTGVGSGKQGS 606
OY 349 AGALKSSNNGRISLLDDVDNEMAIALQGRSMTEQFNNVNPATKELQAMEQLTAM 408
DB 607 IGT-----RVSMILDDAENETASILMSGRQMIHMTENTENPDSQAQOELAAQARA 659
OY 409 SDQLVGADGELPAETQAIKDALAQL--KQPSADGLATAMGOVAFPAAKYVG 458
DB 660 K---AAGDDSAALADAKALEAALGKAGQGGQILNALGQIASAAVVSAG 707

RESULT 4
OYKX33 PRELIMINARY; PRT; 3381 AA.
AC OYKX33;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE StrPA.
OS Streptococcus cristatus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=45634;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCSA;
RA Correia F.F., Allen T.W., Dirienzo J.M.;
RT "High molecular weight serine-rich protein gene (srpA) from
Streptococcus crista";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U96166; AAF34780.1;
DR InterPro; IPR001899; Gram_pos_anchor.

DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 3381 AA; 321845 MW; C623EBE363E74853F CRC64;

Query Match 8.5%; Score 272.5; DB 2; Length 3381;
Best Local Similarity 21.2%; Pred. No. 0.012;
Matches 144; Conservative 130; Mismatches 305; Indels 101; Gaps 19;

OY 9 ESGPVSNSOSNMPIINGQIASNSETKETKA-----EASPSASSVSS 55
DB 1378 QSASASNNONSASISVSASHASISASOSTSASISASISASOSASVSASVSA 1437
OY 56 WFLPSAKNALSLRDAIINKNSPTDSL-----QLEASTSTVTVRAADYDEAKSN 110
DB 1438 SOSTASVSASOSVSASANNONSASISVSASHASISASOSTSA-----SISAS 1487
OY 111 FPTAKSGLENAKTLAEYETKMDLMAALODMERLANSDESNHTEEVNNTKKALEAKOT 170
DB 1488 ISASOSASVSASOSASVSASOSTSASVSASOSASASANNONSA--SISASISASO 1545
OY 171 IDKLKLVTLQNMKSLTEVLTCTDSADQIPAINSOLEINKNSADQIITKLERONIS--YE 229
DB 1546 -----SVASOSASASVSASOSTSASVSASOSVSASANNONSASISVSASHASISASO 1600
OY 230 AVLTNAGEVTKASSEAGIKLQALQSTVDAGDOSQAAVLQAO-----ONNSPDNIA 280
DB 1601 STSASISASISASOSASVSASOSASVSASOSTSASVSASOSASASANNONSASISVS 1660
OY 281 ATKELIDAEYTVNLEKQHTGLTSPLYKKAEBQISOAKDIOETKPSGSDIPYVGS 340
DB 1661 ASOSASISA-----SOSASASISASISASOSASVSASOSASVSASOS--GSA 1707
OY 341 SASASGAAGALKSSNNGRISLLDDVDNEMAIALQGRSMTEQFNNVNPATKELQ 400
DB 1708 SVASOSASASANNONSASISVA---SHASIS--ASOSASASISASISASOSAS 1759
OY 401 MEAOLTVMSDQLVGADGELPAETQAIKDALQKPSADGLATAMGOVA-----FAAA 454
DB 1760 VSASOSASAS--VSASOSTSASVSASOSASASANNONSASISVSASOSASISASO 1817
OY 455 KYGGG--SAGTACTVQMNQVLYKTPAFSTSSSYAALSDGYSAKYTLNLSYSESGVQ 513
DB 1818 SISASISASOSASVSASOSASVSASOSTSASVSASOSASASANNONSA--SIVSAQ 1876
OY 514 SA---ISOTANALRSVRSRGIESGRSADSORAETIVVDSOTLGVYRLYLDEL 570
DB 1877 SASISASOSTSASVSASOSASVSASOSASVSASOSTSASVSA-----SOS 1931
OY 571 MSTIVSNPQ-----ANQEIIMOKLTASISKAPQFGYPAYONSADSLQKA 615
DB 1932 ASASASNNONSASISVSASHASISASOSTSASISASISASOSASVSASOSASVSA--A 1989
OY 616 AOLREFVDCERSLAESQEN 635
DB 1990 SOSTSASVSASOSVSASASN 2009

RESULT 5
OYVQ99 PRELIMINARY; PRT; 2283 AA.
AC OYVQ99;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Serine-threonine rich antigen.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcaceae;
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FR1326;
RA Sharp L.J., Henderson B., Poole S., Nair S.;
RT "Identification of a putative serine-threonine rich antigen from

RT Staphylococcus aureus.":
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF459093; AAL58470.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
SEQUENCE 2283 AA; 228866 MW; 9C0991E0E5B24B0 CRC64;

Query Match 8.5%; Score 272; DB 2; Length 2283;
Best Local Similarity 20.4%; Pred. No. 0.0075;
Matches 144; Conservative 142; Mismatches 308; Indels 112; Gaps 20;

QY 10 SCPESSVSSQSSNNPIINQIANSNETKSTKASESPSASSVS---SKSFSSAKKML 66
DB 814 STSESSSVSSSTSTSLVNSQSVSSMSDASKSTSLSDSTISNSSSTREKSESLSTSTSDSI 873
QY 67 ---ISRDAILKNNSSPTDLSQLEA-STSTSTVTVAAKDYEAKNPDPAKSGLENAK 122
DB 874 RTSTSLSDSL---SMSTSSLSKQSLSTSDSASTSOSVSDSTNSISTAEISLESAS 930
QY 123 TLAETKKNADLMAALQDMERLANSNDPSNNHTEVNNIKKALEAQKDTIDKLKLTVLON 182
DB 931 T-----SDSISISNS-----IANSQASATSKDSQSTSLSTSDS 966
QY 183 QNKSLETVL---KTTDSADQIPAINSOLEINKNSADQIIKIERONISTEAVLTNAGEV 238
DB 967 KSMSTSESLSDSTSTSDSVSGSLVAGSOSVSTSTSDSM-----STSEI 1010
QY 239 IKASSAGIKIKLOALOSIYDAGQSOAAVLQAOQNNSPNIAATRELIDAEKKNVLELQ 298
DB 1011 VSDSITSGSL-----SASDSKSMVSSSMSTSGSTSESLSDSOSTSDSDSKSLST 1065
QY 299 EHTLTD-----SPLVKKAEQ-----ISOAKDIOEIKPSGSDIPIVPSGSAAGASA 348
DB 1066 SGGSGSTSTSTSSVRTSESGSTSGSMSTSGSDSTISSTFSDPTSDSKSAATASESI 1125
QY 349 AGALKSNNSGRI--SLLDVDNEMAAIALQGRSM-----IQGFVNNPATKKEIQ 399
DB 1126 SOSV--STSTSGSVSTSTSLSTSNSESTSTSMSTSLSTSESDSTSTSDISEAIS 1184
QY 400 AMEAQLTASMDQVLGADGELPAETQAIKDALQALKOPASDGLA----- 443
DB 1185 GSESTISISENSSTGDSSEKSAFLSESLSESTSESLSGSTSDSTSLSDSNSES 1244
QY 444 -----TAMQVAPAAKVGGSAGTAGTVQNNVKOLYKTAFTSSSSYAALSDG 494
DB 1245 GSTSTSLNSTSGSTSTSTSGASTSTYKSESVSTSLSTSTSLSDSTSLSTSLSDS 1304
QY 495 YSAVKTLLNLSIESKGVOSAIISQITANPALSRVSRGIESGGRADASORAAETTVRDS 554
DB 1305 TSGSKS--NSL--SASMS--TSDSISTRKSESLASTSLSGSTSESGSTSSASOS--DS 1358
QY 555 QRTGVDYSRLOYDLMSITVSNPOANOEIMQKLFASISKAPOQFPAVONASDLSQKR 614
DB 1359 TSMSS--LSMSQSTISGSTSTSTSLSDSTSLISLASMSQGVDSNSASQASSTSTSI 1416
QY 615 AAOLEREFVDEGRSLAESQENAFRKQPAFIQVLYNIAISLFGSYLS 660
DB 1417 TSSDSDQSTSTYTSQSTSGSESTSTSTSLSDSTSIKSTSGSGSTS 1462

RESULT 6
Q990Y4 PRELIMINARY: PRT; 2271 AA.
AC Q990Y4
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Hypothetical protein SAV2654.
GN SAV2654 OR SA2447.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;

OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21111952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiratazu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003366; BAB58816.1; -
DR EMBL: AP003138; BAB43752.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor.1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SEQUENCE 2271 AA; 227844 MW; 7C2A7040D6C8289D CRC64;

Query Match 8.4%; Score 268; DB 16; Length 2271;
Best Local Similarity 20.6%; Pred. No. 0.011;
Matches 146; Conservative 132; Mismatches 271; Indels 160; Gaps 22;

QY 10 SCPESSVSSQSSNNPIINQIANSNETKSTKASESPSASSVSMSPLSSAKNMLISL 69
DB 1134 SGGSVSTSTSLSTSNSESTSTSVSDSTSLSTSESDSTSESTSTSDISEAISSESTSI- 1192
QY 70 RDAILKNNSSPTDLSQ-----LEASTSTSTVTVAAKDYEAKNPFPAKSGLEN 120
DB 1193 ---LSESNSTDBESQASAFLESELSESTSESVSSSTSESLSDSTSEEG-ST 1247
QY 121 AKTLAEYETKMDLMAALQ-----DMERLANSNDPSNNHTEVNNIKKALEAQKDTID 172
DB 1248 STSLNSTSGASISTSTSTSESTSTFKSESVSTSLSMSTSLSTSLSTSLSDSTSD 1307
QY 173 KLNKLVTLNQNKSITLFEVTKTDS-----ADQIPAINSOLEINKNSADQIIKPL 221
DB 1308 -----SKSDSLSTSMSTSDSTISKSDSISTSTSLSGSTSESDSTSSSEKSDS 1358
QY 222 ERONISYE-----AVLTNAGEVTKASSEAGIKIGQAL-QSIYDAGQSOAAVLQAOQNN 275
DB 1359 TSMSTSMSTSGSTSTSTSTSLSDSTSLISLASMNQGVDSNSASQASSTSTSI 1418
QY 276 PDNIAATKRELIDAEKKNVLEKQEHGTGLDPSLVKKAEEQI-----SOAQKDIQ 324
DB 1419 ESDSQST-STYTSQSTSGSESTSTSTSLSDSTSIKSTSGSGSTSTASISLSESDSQ 1477
QY 325 EIKPSGSDIPIVPSGSAAGSAGALKSSNNSGRIISLLDQVDMEMAAIALQGRSMI 384
DB 1478 STTSSASE-----STSEASTSLSDSTSTSNSSGAS----- 1508
QY 385 EGFVNNPATKELQAMEAQLTAMSDQVLGADGELPAETQAIKDALQALKOPASDGLAT 444
DB 1509 TSTSLSNASASAE---SDSSSTSLSDSTASMSQSSSDSQSTASLSDSLSTSNRMET 1565
QY 445 AMQVAPAAKVGGSAGTAGTVQNNVKOLYKTAFTSSSSYAALSDGYSAVKTLLNL 504
DB 1566 ---IASLSTSVSTSESGS-----TSESTSESDSTSTSLSDSQSTSR----- 1603
QY 505 YSESRSGVOSAIISQITANPALSRVSRGIESGGRADASO-----RAAETIYRDSQTLQD 559
DB 1604 -STSSAGSASTSTSD---SRSTASSTSMRTSTSDSQMSLSLSTSTSMKSDTSLSD 1659
QY 560 VVS-----RIQVLDLS-----IMSTIVSNPOANOEIMQKLTAST 593
DB 1660 SVSDSTSDSTSNSTSGSMVSTSLSDSTSTSTASVEMASISIDSGSMSESVNDSVSE 1719
QY 594 SKA-----PQFGYPAVONS-----ADSLQFAAQLEREFVDEGRSLAES 632

Db 1720 SENSESKMSGSTSVSDGSLVSTLRKSESVSSSSLSGSGMSDS 1768

RESULT 7

099054 ID 099054 PRELIMINARY; PRT; 6713 AA.
AC 099054;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Ebha protein.
GN EBHA OR SA1267.
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OC NCBI_TaxId=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanemori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hiraekawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shibata T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003133; BAB42527.1; -.
KW Complete proteome.
SQ SEQUENCE 6713 AA; 722339 MW; AF6FEDE226BE488 CRC64;

Query Match 8.1%; Score 261; DB 16; Length 6713;

Best Local Similarity 20.5%; Pred. No. 0.078;

Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

OY 10 SGPESVSSNOSMNPILINGQIASNSETKSTKASESPSASSVSSWSFLSSAKNALISL 69
DB 5610 NEDQNLANKKDKANAFVNSLNGLNQOQ-----DLAKHAINNADVPSVTDIVNNQIDL 5663
OY 70 RDAIINKNSPTDSLSQL---EASTSTSTVTVRAAKDYDEAKSNEDTAKSGENAKTLAE 126
DB 5664 NDAM-----ETLKLHVDNEIPNAEQVYVYQND--DNAKTNFDDAK----- 5702
OY 127 YETKADLMAALQDMERLANSPPSNHNEEVNNIKKALEAQKDTIDKLN-----KLV 178
DB 5703 -----RLANTLLNSDNT--NVNDINGAIOAVNDAIHNLNGDQRLQADK 5745
OY 179 TLONONKSLTEVLTCTDSA-----DOIPAINSOLE-----INKNSADQIIKDLERO- 224
DB 5746 AIOGINQALANKLKEIEASNATDOKLAKNKAEEELANSIINNINKATSNQAVSOVQTAG 5805
OY 225 NISYFAVLNAGEVIKASSEAGIKLGOALOSIVDAGDOSQAVALQAOQNSPDNTAAATKE 284
DB 5806 NHAIEQV--HANEIPRAKIDAKKDVQVQALIDEIDR-----NPNLTDK 5848
OY 285 LIDAEETVNE-LKQEHNGLTDS---PLVKAEEQISOAKDIOEIKPSSGDIPIVGPSSG 340
DB 5849 EKQALKDRINQILOGGHNDINNALKKEIEQAQALQALQIDIKDKVAKED----- 5900
OY 341 SNAASGSAAGALKSSNSGRISLLDDVDN--EMAAIALQGRSMIEQF-----NVNNP 392
DB 5901 -----AKODVQKQVQALIDEIDQNPDLTDEKQALKDRINQILOGGHNGINNA 5948
OY 393 ATAKEIQAIEAQLT-----AMSDOLVGADGELPAETQATK 427
DB 5949 MKKEIEQAQKQALQALKEIKLVAKAKENAKODVQVQALIDEIDQNPDLTDEKQALK 6008
OY 428 DALAQALQPSADGLATAGQVAFAPAAKVGSGSAGTACTGVQNVKQLYKTAFTSSSSSY 487
DB 6009 DRINQILOGGHND--INNATKKEIEQAK-----AQALQALQ--DIDKLVK----- 6050

OY 488 AALSDGYAVYKTLNSVSESRSGVQSAISQTPANPALSRVSRSGIESGCRSADASORAA 547
DB 6051 --AKEDAKNAIK--ALANAKRQIINS-----NPDLPPEQAKAL-----KEIDEAKNA 6095

OY 548 ETIVRDSQTLGVDYSRLQV-LDSLMTSTIV-----SNPQANO-----EETM----- 586

DB 6096 LQNVENAQPIIDQNLNGNLGLDIDIRNTHWEEVDQPAVNEIEATPEQIILVNGELIVHRD 6155

OY 587 -----OKLTASISKAPQFC-----YPAVONSADSLQKFAAQLEREFVDGERSLAESQENA 636

DB 6156 DITTEODILAHINLIDQLSAEVIDTPSTATISDSL---TAKVEVTLIDGSKVIVNVPKV 6212

OY 637 FRKOPAFIQO 646

DB 6213 VEKELSVYKQ 6222

RESULT 8

093186 ID 093186 PRELIMINARY; PRT; 6713 AA.
AC 093186;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein ebha.
GN EBHA OR SAV1434.
OS Staphylococcus aureus (strain Mu50 / ATCC 700659).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OC NCBI_TaxId=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanemori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraekawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shibata T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.,
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003362; BAB57596.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 6713 AA; 722306 MW; BBCC0536AC341BF5 CRC64;

Query Match 8.1%; Score 261; DB 16; Length 6713;

Best Local Similarity 20.5%; Pred. No. 0.078;

Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

OY 10 SGPESVSSNOSMNPILINGQIASNSETKSTKASESPSASSVSSWSFLSSAKNALISL 69
DB 5610 NEDQNLANKKDKANAFVNSLNGLNQOQ-----DLAKHAINNADVPSVTDIVNNQIDL 5663
OY 70 RDAIINKNSPTDSLSQL---EASTSTSTVTVRAAKDYDEAKSNEDTAKSGENAKTLAE 126
DB 5664 NDAM-----ETLKLHVDNEIPNAEQVYVYQND--DNAKTNFDDAK----- 5702
OY 127 YETKADLMAALQDMERLANSPPSNHNEEVNNIKKALEAQKDTIDKLN-----KLV 178
DB 5703 -----RLANTLLNSDNT--NVNDINGAIOAVNDAIHNLNGDQRLQADK 5745
OY 179 TLONONKSLTEVLTCTDSA-----DOIPAINSOLE-----INKNSADQIIKDLERO- 224
DB 5746 AIOGINQALANKLKEIEASNATDOKLAKNKAEEELANSIINNINKATSNQAVSOVQTAG 5805
OY 225 NISYFAVLNAGEVIKASSEAGIKLGOALOSIVDAGDOSQAVALQAOQNSPDNTAAATKE 284
DB 5806 NHAIEQV--HANEIPRAKIDAKKDVQVQALIDEIDR-----NPNLTDK 5848
OY 285 LIDAEETVNE-LKQEHNGLTDS---PLVKAEEQISOAKDIOEIKPSSGDIPIVGPSSG 340

```

Db 5849 EKOALKDRINLOOQHNDINNLTKEEIOAAQALQALQDIDKDKAKED----- 5900
QY 341 SAASAGSAGALKSSNNRSISLLDDVDN--EMAIALOGFRSMIEOF-----WNVNP 392
Db 5901 -----AKQVDKQVQALIDEIDONPNULTKEKQALKDRINLOOQHNSINNA 5948
QY 393 AFAKELQAMEAOLT-----AMSDOLVAGDGLPAEIOAIR 427
Db 5949 MKREEIOAKQALQALKEIKDLVKAENAKQVDKQVQALIDEIDONPNULTKEKQALK 6008
QY 428 DALAQLKQPSADGLATAMGQVAFAAKYGGGAGTAGIVQMNVKOLYKTAFSTSSSY 487
Db 6009 DRINLOOQHND-INNAATKEIEQAK-----AQLAQLAQ-DIKDLVK----- 6050
QY 488 AALSDGYAYKTLNLSYSESRGVQSAISOATNPALSRVSRSIGESQSDASORAA 547
Db 6051 --AKEBAKNAIK--ALANAKKQDINS-----NPDLTEPQKAKAL-----KEIDEAKRA 6095
QY 548 ETTVRDSQTLGDVYSRLQV-LDSLMSTIV---SNPOANO-----EEM- 586
Db 6096 LQNVENAOPTIDOLNRLGTLNGLDIDIRNTHWEVDEQPAVNEIFEATPEQILVNGELIVHRD 6155
QY 587 -----OKLASTSKARQFG-----YPAVONSADSLQKFAQLEREVDGERSLAESQENA 636
Db 6156 DITBEDDILAHNLIDQLSAEVIDPPTATISDSL--TAKVEVTLLDGSKVIVANPVKY 6212
QY 637 FRKOPAFIOQ 646
Db 6213 VEKELSVKQ 6222

```

RESULT 9
Q9BITO PRELIMINARY: PRT: 2016 AA.

```

AC Q9BITO:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Fibroin 3 (Fragment).
OS Plectreurys tritaxis (Spider).
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
OC Araneomorphae; Haplogynae; Plectreuridae; Plectreurys.
OX NCBI_TaxID=33319;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-21179804; PubMed-11283372;
RA Gately J., Hayashi C., Motriuk D., Woods J., Lewis R.;
RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
RT Fibroin Sequences.";
RT Science 291:2603-2605 (2001).
DR EMBL: AF350283; AAK30612.1; -.
FT NON_TER
SQ SEQUENCE 2016 AA; 196162 MW; F53AA7DB185826F1 CRC64;

```

Query Match 7.4%; Score 237.5; DB 5; Length 2016;
Best local Similarity 22.2%; Pred. No. 0.15;
Matches 157; Conservative 110; Mismatches 316; Indels 125; Gaps 24;

```

QY 14 SVSSNGSSNNPITNGOIASNSETKSTKASEAPSS-----ASSSV 53
Db 648 STSIAKQTAASTSASTATSTTQTAATTSASTAASQTVOKASTSSASTAAQOTGSSSV 707
QY 54 SSWFSFSSAKNALISLDAIILNKNSPTDLSQLEASTST--STVTRVAKQYDEKSNF 111
Db 708 QNQG-QSSASSSSVSDI-----SDSLTSLQSEFTSAFGTVEAEAOSTAEVAOS 762
QY 112 DPAKSGLENAKTLAEYETKADLMALQDMERLANSPPSNHTEEVNNIKKALEAQDIT 171
Db 763 TVALQGI-----DYSQSSALATAVANAVSQYKQSSSAVYA-----RAIAYVITTY 808
QY 172 DKLNKLVY--LQONQKSLTEVL-----KTDSADQIPAINSOLEINKNSADQIITKDE 222

```

```

Db 809 LKTRITITTTTQVKSFPASAISSLSLTPARNTSSANAYEOTTQSSAAASAAQ--SSEYQ 867
QY 223 RQNISEAVLNTNAGEVIKASSEAGIKQLQALQSIDVAGQSOQAVALQAOQNSPDNIAT 282
Db 868 TQNTQSSA-----SASSDASTSY-QTQDSYDASAASVAA-----ESTSAQAS 912
QY 283 KELIDAEFKVNEKQEHGTGLDPSLVYKKAEEQISOAQKDI----- 323
Db 913 TQSSAAASSSTNSAVQSOOSYIDASTVSSASANTAOFTVQVITPDNTYFAESTLSTLQ 972
QY 324 -QEIKPSSGDIPIVGPSSGSAASGSAAGALKSSNNNG-----RISLLDDVNEMAATA 376
Db 973 EOPNSKFGSTIPLVTAREYASAMARATALLIGFDSGTGSALESAAVAAVSNVDYASISY 1032
QY 377 LOGFRSMIEQFNVNP--ATAKE-----LQAMEAOLPMSDOLVAGDGLPAEIOAIRKAL 430
Db 1033 ARAIARAIISVNLNNQIFPASFASALYLPARMLASLHAFKSSP--SSSAFALANSTISPT 1091
QY 431 AQALKQPSADGLATAMGQVAFAAKYGGGAGTAGIVQMNVKOLYK-----TAFSSTS 483
Db 1092 AITSAOSSSVSAGASSGOSYDTSSVSSASSSEATESSSVFDTYQATQIESSAAAAAAS 1151
QY 484 SSSYAAALSDGYSAKYTLNLSYSESRGVQSAISOATNPALSRVSRSIGIESQ--GRSAD 541
Db 1152 SSAVDSQFSESSSASSAASAFSEOTSYSIDSLSSASTTAASASSAYESQFSDASSG 1211
QY 542 ASQRAETIVRDSQTLGDVYSRLQVLDLSLMSTIVSNPOANOEBIMOKLTAS----ISKAP 597
Db 1212 SSSAAASSSQNSQNDYDIALYSA-----SSAAASAAASAVLEFSDASSSSAAVAAVSSQ 1266
QY 598 QFCYPAVQ--NSADSLQKRAQ--LEREPVDGERS-----LAESQENAF 637
Db 1267 QGSDYDTSDFSSSASSAAAAAASAYESKFLDASSSSSSAAAAAASSQOSSY 1314

```

RESULT 10
O64046 PRELIMINARY: PRT: 2285 AA.

```

AC O64046:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative transglycosylase.
DE YOMT.
OS Bacteriophage SPBc2.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX NCBI_TaxID=66797;
RN [1]
RP SEQUENCE FROM N.A.
RA Lazarevic V., Duesterhoeft A., Soldo B., Hilbert H., Manuel C.,
RA Karamata D.;
RT "The complete nucleotide sequence of the Bacillus subtilis spBc2
RT prophage.";
RT Submitted (AUG-1997) to the EMBL/Genbank/DBD databases.
RL EMBL: AF020713; AAC13005.1; -.
DR HSSP: P00718; 153L.
DR InterPro: IPR002886; Peptidase_M37.
DR InterPro: IPR00189; SLT domain.
DR pfam: PF01551; Peptidase_M37; 1.
DR PROSITE: PS00922; TRANSGLYCOSYLASE; 1.
SQ SEQUENCE 2285 AA; 252310 MW; FF602C227754B357 CRC64;

```

Query Match 7.3%; Score 235; DB 9; Length 2285;
Best local Similarity 19.3%; Pred. No. 0.22;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

```

QY 10 GGPESVSSNQSSNNPIT--NGOIASNSE-----TKESTKASEASPSSASSVSSWFLSS 61
Db 56 SAIDTYQKMLKSYNGVTKETSYTKNADGSVEKLTQGYKKNGEITLQRETKRIINN----- 109
QY 62 AKNALISLDAIILNKNSPTDLSQLEASTSTSTVTRVAKQYDEKSNFD-----TKK 115

```

Db	110	-RNTALKEPTEOVNKLTQATEKLCQVQKTYVRNLQOGPTKVYQAKRNGEDDIYTTTDPK	168
Qy	116	SGLENAKTLAEYETKRMADJMALQDMERL-----ANSDPNSNHTHEEVN	158
Db	169	TNSTSSKTTTNYNDQORRAIEBQLKQDLEKLRQGIYVTDPTTISLGRKINTAQSQAQIEALQ	228
Qy	159	NKKKALEQKOTIDKLNKL-----VTLQONOKSLTEYLKTTDS---ADQIPAI	203
Db	229	NRIRMLDOKSAAYVAKNNELKRTIELYORQAQVNYQNINTRYGSSMGSSNRQAVODYLNAY	288
Qy	204	NSOLEINKNSADQIHKDLERONISYEAULTNAGEVIRKASSPAGIKIGLOALQSI---VDAG	260
Db	269	NS-LWVSTGS-NNTRSQIQSLNMQPRLASQAQTAAQANASSFGELLTQTKRSMSTYISG	346
Qy	261	D-----QSAAYVLAQAOONNSPNDIATAKELIDAEFTYVNLKQOEHTGIDTSP	308
Db	347	SLFYGALSGLEKEMYSQAIEIDTLMTN-----IRRWNNEDYKYNNLLQESIDLGOT-L	398
Qy	309	VKKA-----EEQISOAKODIQEIKPESGSDIPIYPSGSAASAGSAGAL	352
Db	399	SNKITDILQMTGDFGRMGFDESELTSTLTKTAQVLO-----NVSDLTPDDTYVNTLTAAMLNF	454
Qy	353	K-SSNSNGSRISLDDUDVONEMAAIALQ-----GFSMT-----	384
Db	455	NIAANDSISIADKLNEVDNNYAVTTLDLANSIRKAGSTASTFGVELNDLIGYTTAIASTY	514
Qy	385	-EOFWN-----NMPATAKELQAMEAQLTAMDOLVAGDELPAEIOATIKDAL	430
Db	515	RESGIVYVNSLKTIFARIGNQSSIKALEQIGISYKTRAGGEAKSA-SDLISEVAKWMDL	573
Qy	431	AQALKQPSADGLA-----TAM---GOVAFAAKVGGSGAGT-----AGYOMNV	471
Db	574	SDAQKONTSISIVAGIYOLSRFNAMNNMNSIQAONAKTAAANSTGSAWSEQKAYDSIQARV	633
Qy	472	KQLYK--TAFESTSSSSYAAALSGYSAV-KTLNSLYSEBSGYQS-----AISQTA	520
Db	634	NKLONNTEFAIASDAF---ISDGLTEFQOAGSLINASTGVYKSGVLPPLIAAVSTA	690
Qy	521	NPALSR-----SVYRSQIESQGRSADASORAAETIYRD---SQTIG	558
Db	651	TLLSKNTRTLASSLILCTRAMGOETTLATAGLEGMTAPAAVASVYKLTALRGLLVSTLVG	750
Qy	559	DVYSRL-OVLDLSLSTVYINPQANOQEIOMKLTASISKAPQFGYPAYONSADSLQFPAQ	617
Db	751	GAFALGALWALESTLSSPFAKKAKDD-----PQSQOTNVEALTTKKQSDIKLQQ	801
Qy	618	L-EREFVUGERSLAESQENAFRRKQPAFIQOYLVNINIASLFCGY	658
Db	802	YKELQKVESRSILTSDEOEQIQC---TQOILAQCFEPPLAVGYC	840

```

RESULT 11
031976          PRELIMINARY;      PRT; 2285 AA.
AC      031976
AC      031976;
DT      01-JAN-1998 (TREMBLrel_05, Created)
DT      01-JAN-1998 (TREMBLrel_05, Last sequence update)
DT      01-JUN-2002 (TREMBLrel_21, Last annotation update)
DE      YOMI protein.
GN      YOMI.
OS      Bacillus subtilis.
OC      Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC      Bacillaceae; Bacillus.
CX      NCBI_TaxID-1423;
RM      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-168;
RX      MEDLINE-98044033; Pubmed-9384377;
RA      Kunitz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA      Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA      Bottaris R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA      Brouillet S., Bruschi C.V., Caldwell B., Capuano N.V., Carter N.M.,
RA      Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,

```

RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
RA Filiz C.M., Fujita Y., Fuma S., Gallizi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamela D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M.,
RA Kunita K., Lapidus A., Lardinois S., Lauber J., Lazerevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Melillo R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Moore D., O'Reilly M., Ogawa K., Ogwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potworlik S., Prescott A.M.,
RA Pesecean E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Riyolta C., Rooha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scifone F.,
RA Sekiguchi J., Sekowska A., Seror S.-J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weitzengger T.,
RA Winters P., Wipit A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
RT *The complete genome sequence of the gram-positive bacterium *Bacillus*
RT subtilis.":
RL Nature 390:249-256(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP
RC STRAIN-168;
RC Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z99115; CAB14053.1; -;
DR HSSP; P00718; 153L.
DR InterPro; IPR002886; peptidase_M37.
DR InterPro; IPR000189; SLT domain.
DR Pfam; PF01551; Peptidase_M37; 1.
DR Pfam; PF01464; SLT; 1.
DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
KW Complete proteome.
KW SEQUENCE 2285 AA: 252310 MW: FFG60C227754B357 CRC64;
SQ

Query Match 7.3%; Score 235; DB 16; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.22;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

[illegible]

QY 353 K-SSNNGRISLLDDVDNEMAAIALO-----GFRSMI----- 384
Db 455 NIAANDSISIDAKLNEVDNNYAVTLLDLANSIRKAGSTAFGEVELNDLIGYTTAIASTP 514
QY 385 -EOPNV-----NNPATAKELOAMEAOALTAMSDOLVGADGELPAETIOAIKDAL 430
Db 515 RESGIVGNLSKTIARIGNNOSSTIKALEFOIGISVKTAGEAKSA-SDLISVAGKMDL 573
QY 431 AQAALQPSADGLA-----TAM--GOVAFPAKVGGSGAGT-----AGTVOMNV 471
Db 574 SDAQONNTSIGVAGIYQLSRFNAMNNFSAIQNAKTAANSVGSAMSEQOKVADSLQANV 633
QY 472 KOLYK--TAFSSTSSSTAALSDGYSAV-KTLNLSYSESRGVS-----AISQTA 520
Db 634 NKLQNNFTEFAIADAF---ISDGLIEPTQAAGSLNASTGVISYVCFPLPLAAVSTA 690
QY 521 NPALSR-----SVSRSGIESQGRSADASORAETIYRP---SQTG 558
Db 691 TLLSKNRTTLASSLLGTGRAMGOETLATAGLEAGCTRANAASRYLKTRALRGILVSTLVG 750
QY 559 DVSRYL-OVLDLSMSTIVSNPOANOELMOKLTJASISKAPQGYPAVONSADSLQKFAAQ 617
Db 751 GAFALGMALLESLSISFAEAKKAKDD-----FEQSQQTNVEAITNKRKSTDKLIQ 801
QY 618 L-EREFVDEGRSLASQENAFKKOPAFIOQVLVNITASLFCG 658
Db 802 YKELQKVESRSLTSDDEOYLQ---VTQQLAQTEPALVKGY 840

RESULT 12

Q9CH86 PRELIMINARY; PRT; 1063 AA.

AC Q9CH86;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Unknown protein.
GN Y1HD OR L10852.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bojoltin A., Wincker P., Manger S., Sorokin A., Malarme K.,
RT Weissenbach J., Ehrlich S.D., Sorokin O., Malmgren K.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403".
RL Genome Res. 11:731-753(2001).
DR EMBL; AE006319; AAK04950.1; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PRINTS; PR01608; BACINVASINC.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 1063 AA; 104547 MW; 5727ADA9C6969A1 CRC64;

Query Match 7.3%; Score 234; DB 16; Length 1063;
Best Local Similarity 22.4%; Pred. No. 0.094; Indels 160; Gaps 29;
Matches 160; Conservative 108; Mismatches 285;

QY 13 ESVSSNOSMNP1-----INGOIASNSETEKSTKASEAS--PSASSVYSWMSFL-- 59
Db 146 DSSSNNDNSNLSISSNADVDASVQSSTSSGVSALDSGIAVSQSEEML 205
QY 60 -----SSAKNAIISLDAIILNKNSSPTDSLQ-----LEASTSTVTVAAKD-YDEAK 108
Db 206 VGNSSASASAAVASFATLATNPSMVPMLOALAAAPATISGAILNTTTLGDLVNOAI 265
QY 109 SNFPAKSLLENA-KTLAYETK-MADLMAALQDMERLANSNPNHTEVNNIKKALA 166
Db 266 ST--VGISGLANIFSTLGTENIPGMTTAAALNGEVOIVNI-----VGNIOEA--- 311

QY 167 QKDTIDKLKLVTLONOKSLTEVLKTTDSADQIPAINISQLETKNSADQIIKDLERONI 226
Db 312 -----AANPGAFILNETKS-----AGLDVSOQIPLVGGQIAAENAI-----PSM 350
QY 227 SEAVLT--NAGEVTKASSEAGIKLQALQSLVDADQSOAAVLOAQONNSPDNI--AAT 282
Db 351 SPAAMLTLENPPTIPGLSIPGASL--VLSPLVAISVTYSIVN-QLNTTTSNALGVN 407
QY 283 KELIDAEFTKYNELKOEHTGLTDPVKKAEBOISQAOKDIOEIKPSSGDIPIVPGSG-- 340
Db 408 FDLDTLVLSQGDVINYLAAGLVNAINRV-QQIMSQ-----LSPTISNIPVGTIVNN 461
QY 341 -----SAASAGSAAALKSSNNSGRISLLDDVDNEMAAIALQGRSHIEFPNV-- 389
Db 462 VLSPTLNNLTGASLGEVA-----NLGVSSILDQVNNISLGNLISLSTALATIENTLQ 514
QY 390 -----NNPATAKE-----LOAMEAOALTAMSDOLVGADGELPAETIOAIKDALQALKO- 436
Db 515 NSLNFGLNPAGASDILNOVLQONAINNIVESATGIYNNLPG-LGAIENGLSTWISQI 573
QY 437 -----PSADGLATAMGOVA-----FAAKVGGSGAGTAVQOM 469
Db 574 PNINNFVNNALNGITTTIINSLTFPSVGASTVNPNSANSQSSSSASSSSAASSSTSS 633
QY 470 NKKLYKT-----AFSSTSSSTAALSDGYSAKTKTLNLSYSESRGVSQSAISQTANPA 523
Db 634 NVSNVTSSNSSEANTSSSTSNASSSSSESGSSA-----SSSNSESSVASSSSVDSQS 688
QY 524 LRSVSRSGIESQGRSADASORAETIYRPDSQTLDDVYSRLQVLDSMSTIVSNPOANO 583
Db 689 SSAGVNSSSSSAEGSSASSNSSESVASSS-----VSSOS--SSAGVNGS 734
QY 584 EIMOKLTJASISKAPQGYPAVONSADSLQKFAOELREFVDERSLASEQENA 636
Db 735 SSSSSSSASSNSSEGSVAASSSVDSQSSAGVNSSSSSAEGSSASSNSNS 787

RESULT 13

Q9A0N0 PRELIMINARY; PRT; 1086 AA.

AC Q9A0N0;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical protein SPY0697.
GN SPY0697.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacilli/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Plamreau C., Sezate S., Suvoirov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes".
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006522; AAK33654.1; -.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 1086 AA; 115601 MW; 87592EB52CD5ADB CRC64;

Query Match 7.3%; Score 233; DB 16; Length 1086;
Best Local Similarity 21.4%; Pred. No. 0.11; Indels 212; Gaps 34;
Matches 168; Conservative 116; Mismatches 288;

QY 14 SVSSNOSMNP1INGOIASNSETEKES-----TKASEA-----SPSASSSVS 54
Db 355 SMSIGATITIKTYTANQAVSTMTTKADMAKRLSOLGVAKASTIYIYGVGAIISLSTAAITIA 414

RC STRAIN-MH56;
RX MEDLINE-95369882; PubMed-7543881;
RA Jensen L.T., Ladefoged S., Birkeholm S., Christiansen G.;
RT "selection of *Mycoplasma hominis* P321 deletion mutants by cultivation
in the presence of monoclonal antibody 552.";
RL Infect. Immun. 63:3336-3347(1995).
RD EMBL, U21962; AAA81013.1; -
SQ SOURCE 1365 AA: 154983 MW: 8474049924108676 CR64;

Query Match	7.2%	Score 231;	DB 2;	Length 1365;
Best Local Similarity	21.5%	Pred. No. 0.17;		
Matches 149;	Conservative 137;	Mismatches 288;	Indels 120;	Gaps 28

QY	14	SVSSNOSGMNPTINGQJLASNSETSESTKASEAPSSASSVSSMFLSSAKN-----ALIS	68
Db	246	SMQAKSSLDAAV-AETTKKLEFNNKKEAFNKLAKTRNOIOEFINTNNKNPYSLEIS	304
QY	69	LKDAILNKSSPTDLSQLEASTSTSTVTRVAAK-DYDE-----AKSNPDATKSGLENA	121
Db	305	QJTSKRSKSNKSVTSSNKSNDIESANTELKQALAKANADKVOADHLASIKEOJLNNSVNA	364
QY	122	KTL-AEETKKAADLMALQDMER---LANSDPSSNNHTEEVNNIKKALEAQKDTIDKINKL	177
Db	365	NLTLSAKLTKDKNTTQQAQKTELEKEVOKADAIKSNNTASMQSAKSSLDAAVAEITK--KL	422
QY	178	VTLOQNKSLTEVLTQDTSADQIPAINISOLEINKNSAD--QIJKDERONISYEAVL--T	233
Db	423	ETFKKDEAKENELKQTRN-----QIOEFINTNNKNPYSLEISQJTSKRSKSNKSVTSS	477
QY	234	NAGEVIRKASSEAGIKIGQALOSIYDADGQSOAAVLQAQONNSPDNI--AATKELIDAETK	292
Db	478	NKSDIESANTELKQALAKANADKVOADNLAKS--IKEOJLNNSVNAJTLAKLTD---K	531
QY	293	VNELKQEHGTGTDSEPLYKKAQEQISOAQKQIOELFKPGSDIPIYVPGSGSAASAGSAGAL	352
Db	532	DNTTQQAQTEL-----EKEVQKADQAIK-----SNNTASMQSAKSSL	568
QY	353	KSSNNSGRISLLDDVNNEMALIOGFRSMIOF---NYYNPATAKELQAMEAOLITAMS	409
Db	569	DAKVAETTKKLEFNNKKEAFNKLAKTRNOIOEFINTNNKNP---NYSELISQJTSKR	624
QY	410	DQVLG-ADGELPAETQAIKDALQALQKPSADGLATAMGOVAFAAAKVVGGSGAGTACTVO	468
Db	625	DKSNKSVTSSNKSNDIESANTELKQALAKANADKVOA--DNLAKSIKEOJLNNSVNAJTL	682
QY	469	MNV-----KQJYK-----TAFSSSTSSSYAAALSDGSAKTLN--	502
Db	683	AKLTDKNTTQQAQTELEKEJOKRQAKQAIKSNNTASMQSAKSSLDAAVAEITTKKLEFNNK	742
QY	503	-----SLYSESRGCVQASISQTA-NPAISRSVSRSGIESOGRSADASQRAETIVRDSOT	556
Db	743	KEAFNKLAKTRNOIOEFINTNNKNPYSLEISQJ-----ISKRSKSNKSVTSSN	792
QY	557	LGDVYSRLQVLDLSMSTVSNPQANOBEIMOKLTASISKAPQFQYPAVONSADSLQFAA	616
Db	793	KSDIESANTELKQALNTAKAKKSSIDNELPLKNDIOSKEIEFG--PIRNT--NEFWISS	848
QY	617	OLEREYFDGERSLAE--SQENAFKQAFIQOVL	648
Db	849	KLETT---KNKLAEEITKADAIKNNPSSSKQAL	878
RESULT 16			
09KMR3			
ID	09KMR3	PRELIMINARY;	PRT; 2178 AA.
AC	09KMR3;		
DT	01-OCT-2000 (TREMBlrel. 15, Created)		
DT	01-OCT-2000 (TREMBlrel. 15, Last sequence update)		
DT	01-MAR-2002 (TREMBlrel. 20, Last annotation update)		
DE	Streptococcal hemagglutinin.		
OS	HSA.		
OC	Streptococcus gordonii.		
	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		

OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DL1;
RA Takahashi Y., Konishi K., Yoshikawa M.;
RT "Cloning and characterization of the gene encoding a hemagglutinin of
RT Streptococcus gordonii DL1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB029393; BA097453.1; -
DR InterPro; IPR004089; Chm1axis_transd.
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 2178 AA; 203507 MW; 75CC27E27F41DA06 CRC64;

Query Match: 7.2%; Score 230; DB 2; Length 2178;
 Best Local Similarity 19.9%; Pred. No. 0.33;
 Matches 145; Conservative 131; Mismatches 140; Gaps 24

QY	10	SGPSSVNSOSNMPIINGOI-----ASNDETKRSTASSEASPGASSVS-----SWSFL	59
Dd	138	SASBASASANSVSISISISEFSYVASASLSSSSLSOSSSEBASASLSLSVASTSOFS	197
QY	60	SSARNALISLIDALINKNS-----PTSLSQLSEASTSTVTYRAAKDY- -EAKS- - -	109
Dd	198	STSSSTOSSNNESLISSDSSSLTNOSVARNONAVRRIRRAVAAADTAPQKSGDY	257
QY	110	-----NFD-----FAKSGLENAKTLAEY-----TKMAD- -LMAIADMERLANS-----	147
Dd	258	YVRGSEFYEYAEITDNGQVRRVIRVNEGANGSTYLSLPMWVKSTENLPGNATQVNP	317
QY	148	-----DPSNNTHEEVNNIKK-----ALEAKDITI	171
Dd	318	LRTIRFGEVPLNETVNEKSYTTRIIVADMDSGNTQVADNANRGLEFVLTQKSQEKY	377
QY	172	DKLKIYTLQNO-----NKSITFEVLTKT-----SDAQ	199
Dd	378	DPAESSYTYVNNLSNLTSEBREAVALAVRANRPIPTAITIYSONGTVIITYPDKSTDT	437
QY	200	IPA--INSOLEINK--NSADQITIKLERONISYEAVLTNAGEVITKASEAGIKL--GOALQ	254
Dd	438	IPARVYVDLQISKNSASOSSSVASQOSAS-----TSVASISASMASVSVSTASSTS	492
QY	255	SIVAGD--GQAAVLQAQONSPDNITAAKELIDAEFTVNEKQHTLQSLPLAKKA	312
Dd	493	ASVSASEASSTASVSASEASSTASVSASKSSTSTASVSASE-----SASTSASVSAS	546
QY	313	EEQISOAKDIOEIKPGSGSDPIVPGPGSASASAGALKSSNNSGRISLILDDVDNEM	372
Dd	547	ESASTASVSASEASSTASV-----SASTSASTSASVSASEASTASVS- -ASEASTS	600
QY	373	AAILQGRSHITEQFNVNNPA--TKKELQAMPAQITAMS--DOLVAGDELPAEIQIKD	428
Dd	601	ASVSASEASSTASVSASEASSTASVSASESSSTSTASVSASEASTSASVSASEASTS	660
QY	429	ALAAQLQPSADGLATAMGOVAFPAARAVGGSGNACTACTQYMNKQIKTKTFSSSSSYA	488
Dd	661	ASVSASTSASTSASVSASTSASTSASTSASTSASTSASVSASEASTSASVSASEASTS	720
QY	489	AALSDGYSAYTKLNLISXESKRGVQASISOTANALRSVSRSGIEGGRSADASQRAAE	548
Dd	721	ASVSASTS- -STASVSASTSASTSASVSASEASTSASTSASTSASTSASTSASVSASEAST	779
QY	549	TIYVDSQTLGDVYSRLQVLDLSLMTIVSNPDANOELIMOKLTASISKAPOFGYPAVYNSA	608
Dd	780	SASVSASTSASTSASVSASES--ASTSASVSASEASTSASVSASTSASTSASTSASVSASEAS	838
QY	609	DSLQKFAAQ	617
Dd	839	TSASVSASE	847

0938K3
ID 0938K3 PRELIMINARY: PRT: 1086 AA.
AC 0938K3:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Tail protein.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1;
RA Ikebe T., Wada A., Inagaki Y., Sugama K., Tanaka D., Suzuki R.,
RA Katsukawa C., Fujinaga Y., Abe Y., Matsubae H.,
RT Submitted sequence of temperate phage PhiNH1.1.*.
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY050245; AAL15083.1;
SQ SEQUENCE 1086 AA; 115592 MW; 8E9329306629FA82 CRC64;

Query Match 7.1% Score 228; DB 2; Length 1086;
Best Local Similarity 21.0% Pred. No. 0.17; Mismatches 286; Indels 214; Gaps 35;
Matches 165; Conservative 120;

QY 14 SVSSNOSMNPILINGOIASNETKES-----TKASEA-----SPSASSVS 54
DB 355 SMSGATITIKYTVAVQAVSTMTKADMAKRLSOLGVKASVYIVGMGALISLTAATIA 414
QY 55 SMSFLSSAKMALISLDAI-----LNKNSPTDSL-SQLEASTST 93
DB 415 GTAATYALKALVALTGVWVGAIGALVAVGSLMSMLTRKESDETKKKEGGLVES 474
QY 94 STYTRVAAKVDDEAKSNPDITAKSGLENKTLAEYTKMADLMAALQDMERLANSPPSNH 153
DB 475 NKQLR-----DSVEGVOERKKGLESVESTAAHQKLADETIKLAKE-----NKT 520
QY 154 TEEVNIKKALAEOKDTIDKLNKLVTLONONKSLTEVLKTTDSADQIPAINSOLEINK-- 211
DB 521 AGEKRNKLNKIDELNGSIDGLN-LAYDKNSN-----SLSNADQISRSISAMEASTW 572
QY 212 NSADQIIDLERONISYEAULTNAGEVTKA--SSEAGIKLGALOSIVDAGD----- 261
DB 573 QTAQONLINTIQQKRESEKTLAENALRKKNNEANVSDSVKKEIKIETLTEEBAKLKMQ 632
QY 262 ---QSAAVLDAQOONSNDNIAATRE-----LIDAEIKVNEKOE 299
DB 633 TOLDEEYKTSATQOAAADAMAAAEESGASARQVIAYENKSEAOPTAIDNMRTKYSLEJET 692
QY 300 HTGLTDSPLVKKAEOISOAKDIOEIKPS-----GSDIPYVPGSSASASAGSAAALKSS 355
DB 693 TTSIFDA-IEOKTALSIVQOMANLEKNRAATEBOATNIEIL-----AQRGVDSILDL 745
QY 356 NNSG-----RISLLDVNDENMAALALOGFRSMIEQFVNNPATAKELQAMEAQLTJMSD 410
DB 746 RRMPEGATQOVFEVDAIDAEIAPLO-ENFRAATE-----TAK-----NANGS 787
QY 411 QLVGADGLP-----AEIOAIKDA-LAALAKOPSADGLATAMGOVAFAPAA 454
DB 788 VLDSAGVMPKPVKGMVTNVSTGLQAELOANFQOLGOEIPGVQSQISOGAGKASDSASV 847
QY 455 KVGGSAGTACTAGVQMNVKLYTKTAFSSSTSS-----SYAALSDGVS-----AYATL 501
DB 848 KMG-----QEVKRSFOGELIGTSPSRVETFEYGHITTDGLSNVTNCTSKVMOJTM 896
QY 502 NSLYSE-SRSGVQ-----SAISQANPALSRVS-----RSGI-ESGGRSAD 541
DB 897 QSLAQMOSQKQOIVNDMRKSNQITTDATFTMSGPMHSHGVNAGGLANGIYAGGALAA 956
QY 542 ASQRAAEIIVADSOQLGVYSRLOYLDSLMSTIVSNP-----ANOEET---MOKLTA 591
DB 957 AAOSTIAATITATIOSALDIHSPSRVMEDEVGRFI--PGIAGVADIGADRRKVIDDSMOKIKE 1014

QY 592 S--ISKAPQ-----FGYP-----AVONSADSLQKFAALREFEVDGERSIAESQENAFKRP 641
DB 1015 SKTINATPEIASGFGGVAGIANTOTNNNSNFTLVAVDESDE--NSREKQRLFRFES 1072
QY 642 AFIQQ 646
DB 1073 WYIQQ 1077

RESULT 18

0939N5
ID 0939N5 PRELIMINARY: PRT: 3072 AA.
AC 0939N5:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Platelet binding protein GspB.
GN GSPB.
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M99;
RA Bensing B.A., Sullam P.M.;
RT "An accessory sec locus of Streptococcus gordonii is required for export of GspB and for platelet binding."
RL Submitted (Oct-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY028381; AAL13053.1;
DR InterPro: IPR004089; Cmltaxis_transd.
DR InterPro: IPR001899; Gram_pos_anchor.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN.1.
SQ SEQUENCE 3072 AA; 285770 MW; 0B148372657CF7F2 CRC64;

Query Match 7.1% Score 228; DB 2; Length 3072;
Best Local Similarity 20.1% Pred. No. 0.61; Mismatches 310; Indels 88; Gaps 16;
Matches 132; Conservative 127;

QY 10 GSPESVSNOS-----SNMPIINGOIASNETKES-----ASPSASSVSMSFL 59
DB 2426 SASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASTSASTSASTS 2485
QY 60 SAKKALISLDAITLNKNSPTDSLQLEASTSTSTYTRVAAKVDDEAKSNPDITAKSGLE 119
DB 2486 SASESASTS-----ASVSASESASTSAS--VSASTSASTSASVSASESASTSASVSASEST 2541
QY 120 NAKTLAEYETKMAADLMAALQDMERLANSPPSNHTEEVNIIKKALEOKDTIDKLNKLV 179
DB 2542 SASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASEST 2601
QY 180 LONONKSLTEVLKTTDSADQIPAINSOLEINKNSADQIIDLERONISYEAULTNAGEVI 239
DB 2602 SASVSASESASTSASVSASESTSTASVSANESAS-----TASVSASESA 2647
QY 240 KASSSAGIKLGALOSIVDAGD--QSAAVLDAQOONSNDNIAATKELIDAEYTVNEIK 297
DB 2648 STSASVSASESASTSASVSASESASTSASVSASESASTSASVSASESTASVSASE-- 2705
QY 298 QEHGTLTDSPLVKKAEOISOAKDIOEIKPSGSDIPVPGSGA-----ASAGSAG 350
DB 2706 ----SASTSASVSASESASTSASVSASE--SASTSASVSASESASTSASVSASTSASTS 2758
QY 351 ALKSSNNSGRISLLDVNDENMAALALOGFRSMIEQFVNNPATAKELQAMEAQLTJMSD 410
DB 2759 ASVSASESASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASEST 2818
QY 411 QLVGA-----DGLPPEIOAIKDALQALKOPASDGLATAMGOVAFAPAAVGGSGAGTA 464
DB 2819 ASVSASTSASTSASVSASESASTSASVSASESASTSASVSASESASTSASVSASTSASTS 2878
QY 465 GTVOMNVKQLYTKTAFSSSSSYAALSDGYSAKKTLNLSLSESRGQVQSAISQANPAL 524

Db 547 TYOAOI-----QAEIAVQ-----VANVEACEKATAEQEAKKARDE 583
Qy 418 ELPAEIOAIKALAOALQOPSDGLATAMGOVA---FAAAKVGGSACT-ACTYOMNKTQ 473
Db 564 AVKAKKEA-EEKKKQAEAEKTKTATEDANKAKEEAKKASEAKOAEETKAGDVDEEYVA 642
Qy 474 LYKTAFTSSTSSSYAALSDGYSAYKTLNSLYSESRGVSQASISOTANPALSRVSRSQI 533
Db 643 V-NVEFESEVKAARAKAA-----HHKVP EILDKKKNNNAKKAASAKATEKTAEAT 695
Qy 534 E-----SOGRSADAGRA-----AETIVDSQTLGDVYSRLQVLDLSMTIVSNP 578
Db 696 KKATEAKTAAGNAQKASENAKAIADVLAERKASTAQSLKEBAKTL-AADIKKSVTNEE 754
Qy 579 QANOE-----EIMOKLTASISKAPQFGYPAYONSAD-SIQKFAQOLEREVDGERSLAESQ 633
Db 755 KAKRDKAANDAAHQASLSASKAKEKKTAAQAKGVALEKKKEESAKAVEAKKEMKARD 814
Qy 634 ENAFR---KOPAFIOQVLYN 650
Db 815 KAAFEELKTKKODVLEQVDVS 835

RESULT 21
Q9LCJ9 PRELIMINARY: PRT: 1795 AA.
ID 09LCJ9
AC 09LCJ9
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE FmLB.
GN FmTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KS48;
RC MEDLINE=20346625; PubMed=10896508;
RA Komatsuzawa H., Ohta K., Sugai M., Fujiwara T., Giansmann P.,
RA Berger-Bachl B., Suglinka H.,
RT "m551-mediated insertion inactivation of the fnbB gene encoding a
RT cell wall-associated protein abolishes methicillin resistance in
RT Staphylococcus aureus";
RL J. Antimicrob. Chemother. 45:421-431(2000).
DR EMBL; AB015223; BAA93430.1;
DR InterPro; IPR000890; Acetate_kin.
DR TIGRFAMs; TIGR01168; VSIRK_signal; 1.
DR PROSITE; PS01075; ACETATE_KINASE_1; UNKNOWN_1.
SQ SEQUENCE 1795 AA; 191008 MW; 6CB77CCFCB33D350 CRC64;

Query Match 7.1%; Score 226; DB 2; Length 1795;
Best Local Similarity 19.1%; Pred. No. 0.38;
Matches 132; Conservative 129; Mismatches 268; Indels 162; Gaps 25;

Qy 14 SVSSNQSMPDLSQLEASSTSTYTRVAAKDYDEAKSNFDTAKSGLENKTLAEVETKMAAD 73
Db 538 AVEMKKDALQOOVNSQVNSHST-----TASIAEY-----MKLKQOADPTI 577
Qy 74 LNKNSPDSLSQLEASSTSTYTRVAAKDYDEAKSNFDTAKSGLENKTLAEVETKMAAD 133
Db 578 LNKNSPDSLSQLEASSTSTYTRVAAKDYDEAKSNFDTAKSGLENKTLAEVETKMAAD 623
Qy 134 LMAALQDMERLANSDP-----SNHTEEVNIIKALEAQKDTIDKLNKLVTLQONKMS 186
Db 624 KVTAAQSGKYYTQDEVALVTRKINDKNNAIAEIKQTSQVTEKNGIAVLEQD--V 681
Qy 187 LTEVLKTTDSADQIPAI-----NSOLEINKNSAD-----OIINKLE--RONI 226
Db 682 ITTPVKPOAKODIIQAVTTTRKQIKKSNASLQDEKDVANDKIGKIEETKAIKDIDAATNA 741

Qy 227 SYEAVLTNAGEVYIKASSBAGIKLGOALQSIYDAGDOSQAVALQAOONNSPNIATKELI 286
Db 742 QVEAIKTKAINDINOTFTATTAKAALAEFDE-----VQAOIDQALNPDTTNEEV 793
Qy 287 DAEFTKVELKOEHTGLDPSLYKKAERQISOAQKDIOEIKPSGSDPIVPGSGAASAG 346
Db 794 AEAIERINAAKV--SGV-----KAIEATTTAQ-DLERVK-----824
Qy 347 SAAGALKSSNNSGRSLILDDVDNEMAIALQGFMSMEQFN-VNNPATKELQAMEAQL 405
Db 825 -----NEEISKIENITDSTQTKADA-----XNEVQAATVAKRAQNAVYS- 863
Qy 406 TAMSOLVAGDEL-PAEIOAIKALAOALQOPSDGLATAMGOVAFAAAKVGGSAGTA 464
Db 864 NATNEEVALEADAABAQKOGHLDIOVVKSKOEVDTKSKYLDKINAIQTO-----AAVK 918
Qy 465 GTVOMNVKQIKTAFSTSSSYAALSDGYSAYKTLNSLYSESRGVSQASISQTA----- 520
Db 919 PAADTEVENAVYTRKQEOIOTNSV-ASTTEEKQAAAYTELTKKQOARTNLDANNTSAVTTA 977
Qy 521 ---NPALSRVSRSGLIESGSRADASORAE-----TIVRDSQTLGDVYSRLQVLDLSMS 572
Db 978 KDNGIAINQVQAATTKSDAKAEITAKASEKTAIEAMNDSTTEEQAAAEKVQAVLS 1037
Qy 573 -----TIVSNPOANOEIMOKLTASISKAPQFGYPAYONSADSLQKFAQLERE 621
Db 1038 ANADIDNMAANTVDNATTEATTAITTPDANVRQ-----AKQAIADKV-----QAEK 1088
Qy 622 FVDEGERSLAESOENAFKOPAFIOQVLYNIA 652
Db 1089 AIDANNQ-STTEEKAAAKQOVTEKTTADAA 1118

RESULT 22
Q49547 PRELIMINARY: PRT: 1302 AA.
ID Q49547
AC Q49547;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Lmp3 protein.
GN Lmp3.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PG21;
RC MEDLINE=96213016; PubMed=8631664;
RA Ladefoged S.A., Jensen L.T., Brock B., Birkefeld S., Christensen G.,
RT "Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis lmp
RT gene system and identification of gene products";
RL J. Bacteriol. 178:2775-2784(1996).
DR EMBL; X95601; CAA64858.1;
SQ SEQUENCE 1302 AA; 145799 MW; FB9FED1534014A629 CRC64;

Query Match 7.0%; Score 225.5; DB 2; Length 1302;
Best Local Similarity 21.6%; Pred. No. 0.26;
Matches 150; Conservative 139; Mismatches 223; Indels 181; Gaps 38;

Qy 30 IASNETESTKASASPSASSSVSWSFLSSAKKALSLDAILINK-NSPTDLSQLE 88
Db 355 LKSIND-NEISKAESQSLSKRESMESANDLNTK--LLEYRE-LINKRNOEKEAFNELE 410
Qy 89 ASTSTSTYTRVAAKDY--DEAKS--NPDYAKSGLENKTLAEVET--KMAIDMAALQDM 141
Db 411 Q-----TRKNIEFTLDEVKNNPNVATLVKDLTNADAKKSVYNSNKKDIIANBAL 463
Qy 142 ERLANSPPSNHREEVN-NIKKALEAQKDTIDKLNKLVTLQONKSLTEVLKTTDS-- 197
Db 464 IQALADAKKADQVDEANKSIKEQLNA---LIDKANLTPOLINDDS--EIVKAKESINA 518

```
QY 198 -----DQIPAINSQL-EINKN-----SADQIIKLE--- 222
DB 519 EITNANKAVNQNDNASMOQAKSSLDKVTIKQIONLTFEFKDKDAKFEQETRKIDINFL 578
QY 223 -----RQNTISEYAV---LTNAGE---VTKASSAGI-----KLGQALQSIIVADDSQQA 266
DB 579 TDDVKNPNPATLVKDLTNAKDKKSVTKSSNKSEITIAANDELKQAL-----DKAKYA 631
QY 267 VLAQOONNSPDNIAATKELIDAEFTVKNELQEHGTJDSPL-VKKAERQISQAQDIOE 325
DB 632 KQIIDEANK-----SIKEQSDSITMANOLNK-----LVPSDKDIQAKFELS-----QE 677
QY 326 ITPSGSDIPIVPGSGSASAGSAGALKSSNNGRISLLDDVDNEMAAIALOGFRSMTE 385
DB 678 IOSASQELMNNPT-----SMQAKES-----LDAKVTETI-----KKE 712
QY 386 QNNVNNPATKELQAMEAQLTAMSDOLVGADGLP-----AEIOAIKQALQAKOPRA 439
DB 713 TTNKDDVKVFKLEKTRKDI---DEFINTNKTNPYSTLISELTSKRDSKNSVTNSSK 768
QY 440 DELATAMGOVAFAAFAKVGGSAGTAGTVOAMNKOLYKTAFSSSTSSSYAALSDGYSAYK 499
DB 769 SUIETANTELKQALKANKANDKQADNLAKSTKEQLNN-----SISANTLLAKLTD----- 819
QY 500 TINSIYESRSGVQSAISQTPANPALSRVSRGIESQGRSADAS---ORAAETIVRD-- 553
DB 820 -KDNITIOQAKTELEKEV--OKANOAVA-SNNITASMOQAKSSLDKAKVTEIKLETFFKDKD 876
QY 554 -----SOTLGVYSRLQVLDLSMSTIVNPOANDOEIMOKLTASISKAPQCYPAVQNS 607
DB 877 VFKFELQYRKRDI-----DEFINTNKTNP--DYSTLISELTSK-----RDS 915
QY 608 ADSLOKFAAQLEREFEVGERSLAESQENAFKKO 640
DB 916 KNSITNSNKSUIETANTELKQALKANKANDKQ 948

RESULT 23
Q8RJN9 PRELIMINARY; PRT: 1404 AA.
AC Q8RJN9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Variable membrane protein precursor.
GN VMP.
OS Mycoplasma hominis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=132;
RA Boesen T.;
RT "Gene and Protein Structure of the Mycoplasma hominis Vaa adhesin.";
RL Theiss (2001), Department of Department of Medical Microbiology and.
DR EMBL; AJ416752; CAC95143.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN.
FT SEQUENCE 1404 AA; 160137 MW; 2C03F66B3473CAB CRC64;

Query Match 7.0%; Score 225; DB 2; Length 1404;
Best Local Similarity 20.4%; Pred. No. 0.3;
Matches 148; Conservative 123; Mismatches 231; Indels 224; Gaps 31;

QY 83 SLSEASTSTSTVTRVAA-----KDYDEAKSNFDTAKSGL-----ENAKTLAEY----- 127
DB 10 SISLIAAASSTAIAISVCASKNKKFRKQKYNKETEKEQIENAKQAKELKLLLEIKSSDID 69
QY 128 ---ETKMD-----LMAALQDM 141
DB 70 KTNESKVLDSNKIDENSIEDIQNTQIEIAIESLTKRKINDKNKKHEBDOKIVQAMQEF 129
```

```
QY 142 ER-----LANSDDPS-----NNHT-----EENVNINKALEAO- 167
DB 130 KKSQKALGDLINSDQGRVDNSNAKOSLQNNYNNKSSIEQIIQALSKINEAKKELQSOI 189
QY 168 -----KOTIDKLNKLVTLQONKS-LTEVLKTTD--SADQIPAINSQL-EI 209
DB 190 NNARNQKEVEFEEKKOQLIKSNEIDNSKRADEFAILKNNVVGSDISKITETFKTEI 249
QY 210 NK-----NSAQOIIKDLERONISYEAVLTNNGEVTK--ASSEAGIKLQALQSIYDAD 261
DB 250 EKAIESLTNKNINEFKQEQRKANY--KAVFSKSKQKLDLIDSDGK-----VDSEN 299
QY 262 OSQAAY-LQAOONNSPDNIA-ATKEL---IDAETVKNELQEHGTJDSPLVKKAEEQI 316
DB 300 ESQVLTKTIDENSSIEDIQNTKQIEKAIESLTNNKINQKQOKNNLNE--VINNAKELY 357
QY 317 SQAKDIOEIKPSGSDIPIVPGSGSASAGSAGALKSSNNGRISLLDDVDNEKAAYA 376
DB 358 KKLVDSDSEIQOAKTQILDQEIQKASQVVASNDPKAINSSKTS--LDAKITDITKK----- 410
QY 377 LOGFRSMIEQFNVPNNPATKELQAMEAQLTAMSDOLVGADGLPRAIOAIKQALQAKO 436
DB 411 -----LEAFN---ATRK-----LEFTKLOETRSNIDKFLTPREYKA-----N 443
QY 437 PSADGLATAMGOVAFAAFAKVGGSAGTAGTVOAMN--VKOLYKTAFSS-----TSSSYAA 489
DB 444 PNTTLVNL-ELAKKAKEMVSESSNKSQIVAAANNLKAQFQAQSKKADAKTSIEKA 502
QY 490 ALSDGSAYKTLNLSIYSERGVQSAIS-----QTANPALSRV 528
DB 503 KLSASLSNAKKLDKNLTDDGEIQAKAELEVEKANOAITSNNTKEIONSNTSILNKI 562
QY 529 SR-----SGIESQGRSADASORAETIVRDSOTLGDVYSRLQVLDLSM----- 571
DB 563 SEYKKNLDFSEKEKEFEKLEKRSATKEPFINENNTNPYALLOKLOKADAKSTIK 622
QY 572 ---STIVSNPOANDOEIMOKLTASISKAPQCYPAVQNSADSLOKFAAQLEREFEVGER 627
DB 623 SSKSIDIINTQALQALALIEKTERESANSQ--NQAVKWTLNETJAKELDKNLTDSDG 680
QY 628 SLAESQ 633
DB 681 EIQOAK 686

RESULT 24
Q07380 PRELIMINARY; PRT: 1790 AA.
AC Q07380; P89892;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 206.5 kDa protein YDL058W.
GN USO1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloecher H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z74106; CAA98621.1; -.
DR EMBL; Z74105; CAA98620.1; -.
DR SGD; S0002216; USO1.
DR InterPro; IPR002017; Spectrin.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

Query Match 7.0%; Score 224.5; DB 3; Length 1790;
Best Local Similarity 20.9%; Pred. No. 0.43;
Matches 155; Conservative 130; Mismatches 268; Indels 189; Gaps 32;
```

```
QY 3 HHHHMEGPEV-----SSNOSSMN-----PIINGQIASNSE--- 35
Db 969 NNYKQWQAEENSLIKAVEKSESSSIOLNQLKIDMSGOEKNFQIERGSEIKNIEQLK 1028
QY 36 -----TKESTKASEAPSSASSVSSMFLSSAKNAL---ISLDAIINKNSSPPLDSISQLE 88
Db 1029 KTIISDLEQFKREIISKSDSKDEYSQISILKEKLETAFTANDENVNKSISELKTREEL 1088
QY 89 ASTST-----SYTRAAKDYDEAKSNFDTAK-SGLENAKTLAEYEFKMAADLMAAL 139
Db 1089 AELAAKYNLKNLELETLETSKALKVEKENEHEKLEKEATQLEKATQEQOOLSLNLANLE 1148
QY 140 DMERLANSDPNNHTEEVNNIKKALE--AQK-----DTIDKLN-KLVTLQONKSLTEVL 191
Db 1149 SLEK-----EHEDLAAQLKVEEQIANKEROVNEEISQLENDITSTQOENESIKK-- 1198
QY 192 KTTDSADQIPALNS-----QLETKNSADQI--IKLEQNITSEYEVLTNAGEVIRKASSE 244
Db 1199 KNDLEGEVYKAKSTSEBOSNLIKSEIDALNLQIKELKKKNETNEASL---ESIKSVES 1255
QY 245 AGIKLGO-----ALOSIVDAGDQQAVALQAOONNSPDNIATKELIDAEFT 291
Db 1256 ETVKIKLELODECNFEKEVSELEDKLASEDKNSKYLEIQEKSE---KIKELDAKTT 1310
QY 292 KYNELKOEHTGL-----TDSPLVKKAEQISOAKDIOEIKPSGSDIPY 335
Db 1311 ELKIOLEKITTNSKAKKESSELSRLKTSSEERKNAEEQLEKLNKIEIQ-INKQAE--- 1366
QY 336 VPPSGSASAGSAAKALSKSSNGRISLLDDV-----DNEMAAIALOGFRSMIEQFPVN 390
Db 1367 --KERKLNEGST--ITQYSEKINTLEDELIRLQNEENELKAKKIDNTRSELEKVSLS 1421
QY 391 NPA-----TAKELQ---AMEAQITAMSDQVAGDGLPAEIOAKALQALQOPSA 439
Db 1422 NBELLEKONTIKSLDELISLYKDKITRDEKLLSTERNNKRLKLEQLKRAAQ--- 1477
QY 440 DGLATMGQVAPAAKVGGSAGTACTVQMNVKOLYKTAFFSSSTSSVYAALSDGSAYK 499
Db 1478 -----SKAVEEG-----LKKL-----EESSEKKAEELEKSEMMK 1508
QY 500 TINSLSYSESRGVOAISOTANPALSRVSRSGIESQGRSADASQRAETTVRDSQ-TLG 558
Db 1509 KLESTIESNETELKS-----SMETIRKSDKLEQSKKSAEEDIKNLQHEKS 1554
QY 559 DTVSR-----QVUDSLMS--TIVSNPQANOELMOKLTLSISAKPFGYPVONASDSIQ 612
Db 1555 DLISRISEKIDIEELKSKLRIEAKSGSELETYQELNNAQEK---IRINAENTV--LK 1609
QY 613 KFAAOLEREFEVGERSLAESOE 634
Db 1610 SKLEDIERELKDKQAEIKSNQE 1631

RESULT 25
Q9X7M2 PRELIMINARY: PRT: 1327 AA.
AC Q9X7M2:
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DE CTOREF1365, partial (Fragment).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-COL:
RX MEDLINE-99265121: PubMed-10332717:
RA "Mrp a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RL Microb. Drug Resist. 5:9-18(1999).
```

```
DR EMBL: Y09928; CAA71062.1; -.
FT NON_TER 1
FT NON_TER 1327
SQ SEQUENCE 1327 AA; 140517 MW; B90F2085E800586D CRC64;

Query Match
Best local similarity 19.6%; Pred. No. 0.37;
Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;

QY 14 SVSSNOSSNNPIINOIASNSEFTKSTKASEAPSSASSVSSMFLSSAKNMLISLRDAI 73
Db 496 AVEMNKDALQOQVNSQVDSNHYT-----TASIAEY-----NKLKQOQADTI 535
QY 74 LKKNSSPPLDSISQLEASTSTSYTRAAKDYDEAKSNFDTAKSGLENAKTLAEYEFKMA 133
Db 536 LN-----EDANHVTANRASQADIDGLYTKLQAL--IDNQAIAIELDTRKAE 581
QY 134 LMAALQDMERLANSDB-----SNHTEEVNNIKKALEAQKDTIDKLNKLVTLQONKS 186
Db 582 KYTAAQOSKKYQDEVAALVTRKINDKNNAIAEINKQTTAAGVTTREKNGIAVLEQD--V 639
QY 187 LREVLKTTSDAQIPAI-----NSQLEIKNSAD-----QIIRKLE--RONI 226
Db 640 ITPTVYRQAKODIQAIVTRKQOIKKSNASLQDEKDVANDKIGKLETKAIKDIDAATTNA 699
QY 227 SYEAVLTNAGEVIRKASSEAGIKLGOALOSIVDAGDQQAVALQAOONNSPDNIATKELI 286
Db 700 QVEALTKAINDNOTTPATTKAALREFDE-----VQAOIQDQAPLPDPTTNEEV 751
QY 287 DAAETVNLKOEHTGLTDSPLVKKAEQISOAKDIOEIKPSGSDIPVGPSSAASAG 346
Db 752 AEATERINAKV--SGV-----KAIEATTTAAQ-DLEVRK----- 782
QY 347 SAAGALKSSNGSRISLLDDVDNEMAIALOGFRSMIEQFN-VNPNATKELQAMEADL 405
Db 783 -----NEISKIENIDISTQTKMDA-----YEVKQAAATARKQONTVS- 821
QY 406 TAMSQDQVAGDEL-PAEIOATKDALQALQOPSADGLPANGQVAPAAKVGGSAGTA 464
Db 822 NNTNEEVAEDADAQAOGLHDIOVYKSKQEVADTKSKVLKINAITQO-----AKVK 876
QY 465 GTVQNNVKOLYKTAFFSSTSSSYAALSDGSAYKTLNLSYSESRGVO----- 513
Db 877 PAADTEVENAYNTRKQEIQNSN-ASTTEKQAAVTELDTRKQEARFNLDAAANTSDVTTA 935
QY 514 -----SAISQT-----ANPALSRVSRSGIESQGRSADASQRAETTVRDSOTL 557
Db 936 KNSIAAINQVOVAATTKSKDAKAEIQAOKASERTAIEMANDSTTEQQA--KDKVQ 991
QY 558 GQVYSRLQVLDLSMSTIVSNPQANOELMOKLTLSISAKPFGYPVONASDSIQFAAO 617
Db 992 AVVTANADIDNAAANDVNAKTTNATTATATPPDANKP-----AKQAIADKV--Q 1042
QY 618 LREFEVGERSLAESQENAFRQAPFTQOVLVNI 652
Db 1043 AQETALIDGNGG-STTEKKAQAQOVQTEKTTADAA 1076

RESULT 26
Q9RL69 PRELIMINARY: PRT: 2478 AA.
AC Q9RL69:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE MRP protein.
DE MRP protein.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
Staphylococcus.
OX NCBI_TaxId=1280;
RN [1]
RP SEQUENCE FROM N.A.
```

RC STRAIN-COL;
 RX MEDLINE-97302526; PubMed-9158773;
 RA Wu S., de Lencastre H., Sali A., Tomasz A.;
 RT "A phosphoglucanase-like gene essential for the optimal expression
 of methicillin resistance in *Staphylococcus aureus*: molecular cloning
 and DNA sequencing.";
 RL Microb. Drug Resist. 2:277-286(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COL;
 RX MEDLINE-9731478; PubMed-9286983;
 RA Jolly L., Wu S.W., Van Heijenoort J., de Lencastre H.,
 Wenglin-Lecteux D., Tomasz A.;
 RT "The femR35 gene from *Staphylococcus aureus*, the interruption of
 which results in reduced methicillin resistance, encodes a
 phosphoglucosamine mutase.";
 RL J. Bacteriol. 179:5321-5325(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COL;
 RX MEDLINE-99265121; PubMed-10332717;
 RA Wu S., de Lencastre H.;
 RT "Mip-a new auxiliary gene essential for optimal expression of
 methicillin resistance in *Staphylococcus aureus*.";
 RL Microb. Drug Resist. 5:9-18(1999).
 DR EMBL: Y09927; CAB5329.1; -
 DR TIGRFS: TIGR01167; LPTXG_anchor: 1.
 DR TIGRFS: TIGR01168; YSIRK_signal: 1.
 SQ SEQUENCE 2478 AA; 263031 MW; 6B9859A02D023C74 CRC64;

Query Match 6.9%; Score 222; DB 2; Length 2478;
 Best Local Similarity 19.6%; Pred. No. 0.81;
 Matches 136; Conservative 114; Mismatches 275; Indels 170; Gaps 26;

QY 14 SVSSNOSMNPITNGOIASNETKSTKASESPASSSVSSPFLSAKNALISLDAI 73
 DB 535 AVEMNDALQOQVNSODNSHT-----TASTAE-----NKLQQAQDTI 574
 QY 74 LKNSSPDLSOLEASTSTSTVRYAAKDYDEAKSNFPAKSGLENAKTAIEYETKMA 133
 DB 575 LN-----EDANHKTANRASQADIDGLVYTKLQAL--IDNQAIAELDRKAE 620
 QY 134 LMAALDMEFLANSDP-----SNHTEVNNIKALEKQKPTIDKLKLVTLQONKS 186
 DB 621 KVTAAQOSKVTODEVAALVTKINNKNNAIAEINKQTVTEKDNIAVLDEQ--V 678
 QY 187 LFEVLKTTDSADQIPAI-----NSOLEINKNSAD-----QIHKLE--RQNI 226
 DB 679 ITPTVPRQAKODITIOAVTTRKQOIKKSNASLQDEKDVANDKIGKIEFKAIKDIDAATTNA 738
 QY 227 SYEAVLTNAGEVIRKASSEAGIKLQALQSIYDAGDQSOAAVLQAOQNNSPDNTAATKELI 286
 DB 739 QVAETIKTAINDIOTTPTATKAALAEFDE-----VVOAQIDQALVNDTINEEV 790
 QY 287 DAAETKYNELKQHTGLTDSPLVYKKAEEQISOAKQDIOETKPSGSDIPIYVGSASASAG 346
 DB 791 AEAIERINAKV--SGV-----KAIEATTTAO--DLERYK-----821
 QY 347 SAAGALKSSNNSGRISILLDDVDNEMAAIALQFRSMIEFQ--VNNPATAKEILOAMEAQL 405
 DB 822 -----NEISKIENTISTOTKMDA-----YNEKQQAATARKKQONATVS- 860
 QY 406 TAMSQOLVGADGEL-PAEIOAIKDALQALQPSADGLATAMGOVAFAAKVGGSAGTA 464
 DB 861 NATNEVEAEKDAVDAQAQKGLHDIOYVKSQEVADTKSKVLKINAIOQ-----AKYK 915
 QY 465 GTVQNMVKQLYKTAFSSTSSSYAALSDGYSAVKTLNLSYSESRGQV-----513
 DB 916 PADTEVENAYNTRKQEIQSN--ASTEEKQAAYTELDTKQKQARTVLDAANTNSDVTTA 974
 QY 514 -----SAISQT-----ANPALSRVSVSGIESQGRSADAGRAAETIVRSQTL 557
 DB 975 KDNISIAAINVOAATTKKSDAKAEIAOKASERTTAIEAMNDSTTEEOAA-----KQKVDQ 1030

QY 558 GDVYSRLQVLDLSMTSTIVSNPQANOELMOKLTASISAKPQFGYPVQNSADSLQKFAAQ 617
 DB 1031 AVYTANADIDNNAANDVDNAKTNTNATITPAITPDANVPR-----AAKQAIADKV-----Q 1081
 QY 618 LEREYDGERSLAESEQENAFKQPAFIQVLYNIA 652
 DB 1082 AQETALDGNNG--STTEKAAAKQOVOTEKTTADAA 1115

RESULT 27

ID 015738 PRELIMINARY; PRF; 924 AA.
 AC 015738;
 DT 01-JAN-1998 (TREMBLREL. 05, Created)
 DT 01-JAN-1998 (TREMBLREL. 05, last sequence update)
 DT 01-JAN-1998 (TREMBLREL. 05, last annotation update)
 DE Zlpa (Fragment).
 GN Zlpa.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_Taxid=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX4;
 RA Loomis W.F., Trautfar N.;
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF019980; AAB70839.1; -
 FT NON_TER 924
 SQ SEQUENCE 924 AA; 105060 MW; 63921CCA596AB406 CRC64;

Query Match 6.9%; Score 221.5; DB 5; Length 924;
 Best Local Similarity 21.1%; Pred. No. 0.25;
 Matches 148; Conservative 122; Mismatches 282; Indels 149; Gaps 29;

QY 12 PESVSSNOSMNPIT--INQOIASNETKSTKASESPASSSVSSWS-----FLSAKN 64
 DB 278 PESNSPINOSTNDLQOYNTLNGNNNTTTNKKKEITELQSLNISAKEIMVQYQ 337
 QY 65 ALISLDAI---LNKNSPDLISOLEASTSTSTVTRV-----AAKDYDEAKSNFPAKSG 117
 DB 338 QIVITLQOQVSTFEKVDLITNLVSKQE--TKIGELTRANGFTTKTEILRSYEDKKRT 395
 QY 118 LENAKTIAEYE-----TKMADLMAALDMEFLANSDPSSNHTEVNNIKKA 163
 DB 396 AELLERLEMEYKMKNTITEKDEFOIEKIVDQLEAKQSEQOTT--TNNIQNEISOLKQO 451
 QY 164 LEAKQDT-----IDKLNKLVTLQONKSLTEVAKTDS-----ADQIPAINSOLEINKNS 213
 DB 452 LASNSTESQALQSKITELSQLQSEFEKLQNLQSKDSELETSTKRQALREQSEDSQS 511
 QY 214 ADQIHKLE-----RONISYEAVLTNAGEVIRKASSE-----AGIKLQALQS 255
 DB 512 KDEKLKSVELNLQOITLQOLQSKQDELQNVKSQLQSESESESKDQKLKSVELTLOQTLOQ 571
 QY 256 IYDAGQSOQAVALQAOQNNSPDNIATKELIDAETKYNELKQHTGLTDSPLVYKKAEEQ 315
 DB 572 LDQVKSQLEQ--QSEHNESKD-----EKLKIEIENLQOOLQS-----RDSLSK--DEQ 617
 QY 316 ISOAQKDIOEIKPSGSDIPIYVGSASAGSAAGALKSSNNSGRISILLDDV---DNEM 372
 DB 618 LKCLESELSSVKQDLS-----QSSNDSELSVYKQDLSKQDEL 657
 QY 373 AA-----IALQFRSMIEFQVNNPATAKEILOAMEAQLTAMSDQVAGADG---LP 420
 DB 658 KSKDQELSNKDSOIKTIESDQSVKQDLSKQDELQSTKQDLSKQDLSNKTQIKSTIE 717
 QY 421 AETIOAIKDALA-----QALK-----QPSADGLATAMGOVAFAAKVGGSAGTA 464
 DB 718 SDLQSVKQDLSKQDELQSTKQDLSKQDELQSTKQDLSKQDELQSTKQDLSKQDELQSTKQDLS 774
 QY 465 GTVQNMVKQLYKTAFSSTSSSYAALSDGYSAVKTLNLSYSESRGVSQSAISQITNPAL 524

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003615; AAF52458.2;
 DR FlyBase; FBgn0031869; CG18304.
 DR InterPro; IPR002017; Spectrin.
 KW Hypothetical protein.
 SQ SEQUENCE 1833 AA; 206010 MW; ED1948482573584E CRC64;

Query Match 6.9%; Score 219.5; DB 5; Length 1833;
 Best Local Similarity 21.6%; Pred. No. 0.7;
 Matches 145; Conservative 105; Mismatches 265; Indels 155; Gaps 26;

QY 16 SSNOSNMPIIINGQIASNSETKESTKAS-EASPSASSSVSSWSFLSAKNAKILSLRDAIL 74
 DB 250 SSNSS-----TSLSNSTASNAEVKAYVTSTSSSTSSSSSVKREKADSVASKEL 299
 QY 75 NKNSSPTDSLSLEASTSTSTVTRVAAKDYDE--AKSNFTAKSGLENATLAE-YETKM 131
 DB 300 KROTVPAASTSH--SNSTSTASTASKSODINGMOEQMKALTELETMTKTRAKEREK 356
 QY 132 ADMAALQDMERLANSPSNHT--EEVNNIKKALEAKQDTID-----KLN-KLVTTIQ 181
 DB 357 SDLL-----LRLASMDTASNRTAASEBALNLQCKLNEKQDLRVYEDKKRLKRLKELE 411
 QY 182 NQNKSLTEVLTCTDSADQIPAINSOLEINKNSADQIIKDLERONISYEAVLTNAGEVYKA 241
 DB 412 NKS-----ESELRRKLQAEQICEELMERNQSKRKLTL----- 446
 QY 242 SSAGIKLQALQSIYD-AGDQQAAYLQAOQNNSPDNIAATKEL-----IDAETKVN 294
 DB 447 -----LQAEDEVDQDFRDEEVAKAKTSLQCKLEKATKNCRLISFKELKKSRIE 495
 QY 295 ELKQEHGTGLDPL--VKKAEQI-----SOAKKDQIEIK-PGSGDIPITVPGSGSAA 343
 DB 496 TLEQEQSSNNAELSKIKKEELKELFSELTETKLLQAEAEELNPGKKAKAPMLGVLEKST 555
 QY 344 SAGSAAGALKSSNSGRISLLDDVDNEMAAIALQGRSMIEQFNVPNNPA-TAKELQAME 402
 DB 556 SA-DAKFTRESLTRGG-----SQEDPQHQLRELQDST 586
 QY 403 AQITAMSDQLVGADGELPAETQAIKDALQALQKPSADGLATMGQVAAKVGGSAG 462
 DB 587 ERTDTLCKDLKFAEEL--QRLRDRERKRVFSCGTQTEVPLEVAVAFPR-----GTQ 636
 QY 463 TACTVOMNKKQVKTAFSSSTSSSYAALSDGYSAYKTLNLSSESQVQASISQTANP 522
 DB 637 TVATVOSDM-----STSEVENLVTSNVAVTQD--FEVPRNVSIERTWSSPAGLFP 688
 QY 523 ALSRSVRSRGIE-----SQGRSADASQRAETTVRDSQTLGDVYSRL 564
 DB 689 SSSSRVGGSGRKLSPTPHRLAEVHADREGIS--DEDPARLRLTLELNEQEAISLRL 747
 QY 565 QVLDLSMTITVSPQANQOEIEMOKLTASISKAPQFGYPAVONSADSLQFAOLEFEVD 624
 DB 748 KYEDLEKENAESKKYV--RELQAKLRQDSSNGSKSLSLGTSSAAEKVKVLTINEELYQ 805
 QY 625 GERSLAESOE 634
 DB 806 LKRTLEKEQ 815

RESULT 30

007569 PRELIMINARY; PRT; 2139 AA.

AC 007569; 002504;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Myosin heavy chain.
 GN MHCA.
 OS Entamoeba histolytica.

OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_TaxID=5759;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HMI:IMSS;
 RX MEDLINE=93295430; PubMed=8515774;
 RA Raymond-Denise A., Sansonetti P., Guillen N.;
 RT "Identification and characterization of a myosin heavy chain gene
 (mhca) from the human parasitic pathogen *Entamoeba histolytica*.";
 RL Mol. Biochem. Parasitol. 59:123-131(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HMI:IMSS;
 RA Guillen N.;
 RL Submitted (Feb-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; I03534; AAB48065.1; -
 DR HSSP; P08799; 1MD.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR000048; IQ_region.
 DR Pfam; PF00612; IQ_2.
 DR Pfam; PF00063; myosin_head; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00242; MYSC; 1.
 SQ SEQUENCE 2139 AA; 245225 MW; C68307341DB51DD1 CRC64;

Query Match 6.9%; Score 219.5; DB 5; Length 2139;
 Best Local Similarity 20.9%; Pred. No. 0.85;
 Matches 151; Conservative 120; Mismatches 272; Indels 179; Gaps 32;

QY 13 ESYSSNOSNMPIIINGQIASNSETKESTKASEASPSASSSVSSWSF-----LSSAKNAL 66
 DB 994 ESIDEKEDEITYK-LKGDIKLLEEKEDLDQDRADVATKDKLAKKINKITIECEDAKDEI 1052
 QY 67 ISLRDAI--LNKNSPTDSLSQLE--ASTSTVTRVAAKDYDAKSNFTAKSGLEN 120
 DB 1053 AKLEQLEDEENKNDLTJELQTLKGLTEFSLAQVAAT--KRAQSERDTLSQNLN 1110
 QY 121 -----AKTLAEVETKMAADLMAALQDMERLANSPSNHTEEVNNIKKALEAQDTID 172
 DB 1111 EKLTTNLTNRKADLEKRTISGLKQDYEDLE--DDKKIBEDLNAGKRIELEDDET 1165
 QY 173 K-----LNKLVTLQONKSLTEVLTCTDSAD-----QIPAINSQL--EI 209
 DB 1166 KGADVSQYLQKQKEEYESQIAKQKEKALGNQVKNKKEKTIKEKELEIOSLQCKLEDETEY 1225
 QY 210 NKSAAQIIKDL-----ERQNTSYEAVLTNAGEVYKASEAETIKLQALQSIYDVG 260
 DB 1226 EKEDAKKKRKEIEKKAKLQEEKENVESS--KNSTEKDKKKLEDLKDTQCKLDDMTAD 1282
 QY 261 DQ--SQAAVLQAOQNNSPDN--IAATKELID--AAETKVNELKQEHGTGLT----- 304
 DB 1283 NEKTKAKADLEQALNBYQDNHKKAVADAEELLNKKRAQSDKELNSLAKALELTKRKKYV 1342
 QY 305 -----DSPLVYKA-EQIISOAKDIOEIKPSSGDIDIVPSSGASASAGSAGALKSSNNS 358
 DB 1343 ESKNKDSENEKALSEIIOANKEKLNIO--ADL-----RKATADLOEANEKKAEVEBA 1393
 QY 359 GRISLLD-----DVND-----EMAAITLQGRSMIEQFNVPNNPATAELEQ 400
 DB 1394 QRDKLIVADKKKMTKLTETIKARDEENTYVEENEKYLKKREADLEFANENLDIEKKDRN 1453
 QY 401 MEAQLTAMSDQLVGADGELPAETQAIKDALQALQKPSADGLATMGQVAAKVGGS 460
 DB 1454 KEQVAKKLEGELEKTEKDKLNAAI-AEKDSIFPAKKQSDAD----- 1492
 QY 461 AGTAGTVQNNVQVLYTA-----FSSTSSSYAALSDGYSAYKTLNLSSESRSQVQ-- 513
 DB 1493 -----LEELNTKVEHEDEVVAKLNTQITKLRDQNSAEELNELRSADKKKKRI 1542
 QY 514 SALSQTANPALSRVSRSGLSGRSADASQRAETTVRDSQTLGVYSRLQVLDLSMT 573
 DB 1543 SELEQVNELESRPVGTG-----NADEN--EIKIRDAQ-IADLNKALE--MKG 1585

DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1935 AA; 221530 MW; 7C4245EA440B124 CRC64;

Query Match
Best Local Similarity 22.5%; Pred. No. 0.94;

Matches 161; Conservative 101; Mismatches 261; Indels 194; Gaps 33;

```

QY 20 SSMNPIINGOIASNETESTKASPSASSVSSW-----SFLSANKALIT- 67
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 833 NKVRPLN--IARQD--ENKKAQDEFAKMKKEEFASCEOMKRELEQNTVLMOQRNDLVI 888
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 68 ---SLRDAI-----LNKNSPTDSLQLE-----ASTSTVTVRVAAKDYE 106
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 889 AMSSEDAIGDAEEKIEQLIKQKSDFTQIKELDKLMEDEDAATVLSQKKSAEIGE 948
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 107 AKSNDDTAKSGL---ENAKTLAEYETK--MADLMALQMERLANSPPSNHTEEVNNTIK 162
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 949 LKKDVEDLEAGLAKAEQKTKKDNQIKTLQDEMA--QODEHLSTLKKKKLLEVO--KK 1004
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 163 ALE---AOKDTIDKLNLVTLQONKSLTEVLTQDSDAQIPA-----INSOLEIKNSAD 215
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1005 TLEDQAEEDKVNHLSTKTLQEO--TLEDELDNLEREKKIPGDVDKAKKRYEDDLKTTQ 1062
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 216 QIKDLEKQNI SYEAVLTNAGEVTKASSEAGIKLQALQSTIVDAQDQQAAYLQAOQNS 275
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1063 ETVEDELER-----VKRDLQDAGRKKDMKMEINGLNSKLEDBQNLVLAQLQKK-- 1106
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 276 PDNIATATRELIDAAETKYNELKQEHTEGLDPLVYKAEQISOAKDIOEIKPSSDPI 335
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1107 -----IKELQARIELELEELAEERQART-----KVEKQRTSLRELEEL---GERLD- 1150
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 336 VGPSGSAASAGSAGALKSSNNSGRISLLDDVDNEMALQGRSMIEQFNVPNPTA 395
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1151 -----EAGGATPAQOMELKKRQELLRLRLDEEATWQ---HESQATLIRKKNQEAR 1199
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 396 KELQAMEAQLTAMSDQLVAGDELPAETQAIKDALQA-----LKQPSA----- 439
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1200 NELGQIQIQQLQKVSRLKEKTKQLRAEMDVQGVQEHGKNGCKSEKMSKQWAEQLSELN 1259
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 440 ---DGLATAMQVAFMAAKVGGGSGAGTACTV---OMNFKLYKTAFSTSSSYTAALSD 493
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1260 AKIDQASVSSELSQKSRLOTTEADLTROLEAHNVGOLTKL-----KSLIASLSD 1313
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 494 GYSAYKTLN-----SLYSRS-----GVQSAISQTPANPA--LSRSVSSSGTE--SQG 537
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1314 ---AKRSLEDGRRLAKQLQAEVRLNLSIDIGRESLEEEASKSDQLRALSNAAVEQOW 1370
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 538 RSADASQRAETIVRDSOTLGDVYSRLQVLDSLMSTIVSNPOANQEEIMOKLTASISKAP 597
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1371 RSKFSEGAANA-----DEL-----EDAKRKRLQAKLSFAE 1400
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 598 QFGPRAVONSDSLQKFAQLERE-----FVDEGSLAESQENAFKQOPAF 643
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1401 Q-----TADTLHSKAGLEKAKSRLOQETELEDLAIDVERSSAHA--NNLEKKQNF 1448
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|

```

RESULT 35

Q9HGK6 PRELIMINARY; PRT; 2297 AA.
AC Q9HGK6;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE Agglutinin-like protein A1s7p.
GN A1S7.
OS Candida albicans (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1161;
RX MEDLINE=20321177; PubMed=10861907;

RA Hoyer L.L., Hecht J.E.;
RT "The ALS6 and ALS7 genes of Candida albicans.";
RL Yeast 16:847-855(2000)
DR EMBL; AF201684; AAF98068.1; -
SQ SEQUENCE 2297 AA; 244723 MW; 598020C63027651 CRC64;

Query Match
Best Local Similarity 17.9%; Pred. No. 1.2;

Matches 146; Conservative 156; Mismatches 309; Indels 204; Gaps 29;

```

QY 10 GCPES---VSSNQSSMNDIINGOIASN-----SETRKS--TKASEAPSP--ASSSVS 54
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1130 SGMSRHYTNSTETSVDSVSSVAGDETSSESVSISESSSESVASESVASEVA 1189
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 55 SMSFLSSAKNALLSLRDAIILKN-----SSPDSLQLEA-----ST 91
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1190 SEVSASESVTAVIDSLDYTTSEEVSTSDNSGMSPPISSEQRSSIPIMSSDSESSSR 1249
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 92 STSTVTVRVAAKDYEAKSNFDT---AKSGLENAKTLAEYETKMDLMALQMERLANS 148
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1250 ESSSGTILSEENSDSIPPTFTSRWSPSGMSRHYTNSTETSVDSVSSVAGDETS 1309
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 149 PS--NNHTEEVNIIKALE-----AOKDTID--KLKLVTLQONKSLTEVLT 191
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1310 VSVISESSESVTSESVASESVASESVASESVTAVIDSLDYTTSEEVSTSDNSGMS 1369
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 192 KTTSADQIPALINSQLE-----INKNSADQIIKLEQONISYEAV----LTNA 235
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1370 PSSQRSSIPVWSSDESSESSSGTILSEENSDSIPPTFTSRWSPSGMSRHYTN 1429
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 236 GEV-----IKASSEAGIKIGQALQSTIVDAQDQQAAYLQAOQNSPDNIATRELIDAAET 291
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1430 TETSVDSVSSVAGDETSSESVSISESSSESVASESVASESVASEVA-----EVT 1483
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 292 KYNELKQEH-----GLTDSPLVYKAEQISOAKDIOEIKPSSGDIPIVPGSG--SAASA 345
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1484 AVSDIDSLDYTTSEEVSTSDSKIVASTVPSSEQR-----SSIPIMSSDSESSSR 1533
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 346 GSAGALKSSNNSGRI-----SLDDVDNEM 372
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1534 ESSSGTILSEENSDSIPPTFTSRWSPSGMSRHYTNSTETSVDSVSSVAGDETS 1593
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 373 AAILQGRSMIEQFNVPNPTAKELOAME-----AQLTAMSDQLVAGDELPAE-- 422
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1594 VSVISESSESVTSESVASESVASESVASESVTAVIDSLDYTTSEEVSTSDSKIV 1633
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 423 ---IQAIKDALAQA-----LKQPSADGLATAMG-----QV 449
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1654 RSSIPIMSSDSESSSRSSSGTILSEENSDSIPPTFTSRWSPSGMSRHYTNSTETSV 1713
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 450 AFAAKYGGGAGTAGTYOMNVKQLYKTAESSTSSSYA--AALSDGSAKTLNLSLYSE 507
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1714 SDVSSSVAGDETSSESVSISESSSESVTSESVASESVASESVASESVTAVIDSLDYTT 1773
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 508 SRSGVQSAISQTPANPAISRSVRSIGI---ESQGSADASQRAETIV--RDSOTLGDVYS 562
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1774 SEVVSTSDNSGMSPPISSEQRSSIPVWSSNESSESSSGTILSEENSDSIPPTFS 1833
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 563 RLQVLDL-----MSTIVSNP-----QANQEEIMOKLTASISKAPQFGPAY----- 604
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1834 TRYVSVLITVQELSLPQLPKLHLPSSLSLETSGIKTKANLNPQFFSTVDLSALSYMA 1893
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
QY 605 --ONSADSLQKFAQLEREFPYDGBRS--LAESQENA 636
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|
Db 1894 SGSSSADHQSSATCDVSESVQGNLSAMAPGMSNS 1928
   ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:| ||:|:|

```

RESULT 36

083472 PRELIMINARY; PRT; 2055 AA.
AC 083472;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

01-JUN-2002 (TREMblrel. 21, last annotation update)
DE Extracellular matrix binding protein (Fragment).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Aerococcaceae; Abiotrophia.
OX NCBI_TaxID=46123;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NVS-47;
RX MEDLINE=99081722; PubMed=9864195;
RA "Characterization of emb, a gene encoding the major adhesin of
RT Streptococcus defectivus."
RL Infect. Immun. 67:50-56(1999).
DR EMBL: AF067776; AAD0320.1; -
DR TIGRFBAS: TIGR01168; YSIRK_signal. 1.
FT NON_TER 2055
SQ SEQUENCE 2055 AA; 215642 MW; 9699C11DDE933E2FD CRC64;

Query Match 6.8%; Score 216.5; DB 2; Length 2055;
Best Local Similarity 21.3%; Pred. No. 1.1;
Matches 178; Conservative 124; Mismatches 272; Indels 263; Gaps 42;

OY 11 GPEVSSNOSNMPTI-----NGOI-----ASNSEKSTKASEASPSASSVS 54
DB 501 GPELT-----PYVAALRIYPRYNSGOVPEHRTSGNATNAAPTIAVTPPTLLND 552
OY 55 SWSFSSAKN--ALISIRDAIINKSSPT--DLSOLEASTSTSTYTRVA 101
DB 553 QAISDPTIKSMVRVNDLEDADQOTGTPQADIVSYSKNGQVAIDIPVEGEYTVTLA 612
OY 102 KDYEAKSNFTPAKSGLENAKTAEYETKMDLMAADMERLANS--DPSNNHTEEVNN 159
DB 613 RSGQKLSN-NIVQ--VKVAFLLPREAK-----NAVNAAKAKNTAIDONNNLTAEEKA 664
OY 160 IKKA-LEAKDT----IDKLKLVTLQN--QKSLFELVKTDSADQIPAINSOLEINKN 212
DB 665 AKKAYEAARKNKTLAGIDDA-KTTAARNNAOKGTTDINAVPVPVAKPAANAAL--QA 721
OY 213 SADQITKIDLERONISEYEAULTNAGEVIKASEGIRLGOAL--OSIVDGDQSOAAVLA 270
DB 722 AVNKINEISQRPDLTREKQAFMDQVKTARDAAMAKVASAANNQVTSARDGLNAV--- 778
OY 271 QONNSP-----DNIAATKE-----LIDAEETKVN-ELK 297
DB 779 --NNLPTPAKXPALGHVQAADAKRQAIKRONANLTAEEQADALROVDAQTAEAALN 836
OY 298 QHTGLTDSPLYKKAEEQISOAKDIOETKPSGSDIPYVPSGSAASAGSAAALKSSNN 357
DB 837 QHTTNAT--LAKAD--SDGVKAINDINPO--PRSKPANAOLEQVAAAKROAINN 885
OY 358 SGRISLLDDVQVNEA-AI-----ALQGRSMIEQFNVPNPAT----- 394
DB 886 NNQLT-----DEKKAQAIQOYDQALANAKTOYQAANDNNNGVNOAKTACTTAIINNINPOG 939
OY 395 ---AKELQAMEAQULTAMSQOLVG-----ADGELPAETQAIKDALAOLKOPSD 440
DB 940 TQKAQAIATAEAQAKRELQGRNDLTTEERNNNLADLTAAQAQAKQAVNOAARNNTGVA 999
OY 441 GL-----ATAMGOVAFPA 453
DB 1000 GAKDNGVAQIQINPTAVVYPDARNAIDQADARKAEAFQAKNTLDEEKAAMAIKKVQDAA 1059
OY 454 --AKVGGSGAGTAGYQVMVQKLYTAFSSSTSSSYAALLSGY--SAYTILNSISSESS 510
DB 1060 RDAKKAIDRAGSNGDVNNAVNO--GKAALQAIK-----ALDSQPSAKTATAAALIONAD 1112
OY 511 GVQASITAPANALSRVSRSIGI--ESQGRSA--DASQRAET-IYRDS--OTLGV 560
DB 1113 AKKAAI--TANNAULTQEEKAAMAIKQVEDEAQAQAQAVDASRSKADVDRAKKGLOKISIV 1170
OY 561 -----YSRLQVLDLSLMS--TIVSN--POANOEIMOKL----- 589

1171 PAVQPKLNAIAAYDQAAATDKKAVINNDTTLTQEEKEAIIKRVDEEAKARQAINDATSN 1230
OY 590 -----TASISKAPOFGYPAVQNSADSLQKFAAOLEREPFVCGESLAEQENA 636
DB 1231 ADVAAKQAGTOAIVNVPO--TPAAKNKAKAVEQADAAKQAIENDPNLTROEKDA 1285

RESULT 37
O98M03 PRELIMINARY; PRT: 2124 AA.
ID O98M03
AC O98M03
DT 01-OCT-2001 (TREMblrel. 18, Created)
DR 01-OCT-2001 (TREMblrel. 18, last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, last annotation update)
DE Kinesin-like protein.
GN MUR0796.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Ideasa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti."
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002995; BAB48310.1; -
KW Complete proteome.
SQ SEQUENCE 2124 AA; 229253 MW; 3457A2B462EFED01 CRC64;

Query Match 6.7%; Score 215.5; DB 16; Length 2124;
Best Local Similarity 21.6%; Pred. No. 1.2;
Matches 167; Conservative 124; Mismatches 262; Indels 221; Gaps 36;

OY 5 HHMSEGPESVSSNOSNMPTING-----QIASNET----- 36
DB 261 HSEVNOIERSYSENETRIRKSLVDGLGSEAEVYTHAERYRASIAQAEHLRDEIGAASDI 320
OY 37 -----KESTKAS-----EASPSASSVS-----WSFLSAKNALLSLR 70
DB 321 IYDSTLNASTKLSMTTNSGDTLIDRINNESSMSIFDSVGRDITTDKLISTGGEAFASLL 380
OY 71 DALINKNSPTD---SLSOLEASTSTSTVTRV--AAQDYDEAKSNFTPAKSGLENAKTL 124
DB 381 DTRIAKLTFTDGLTFRSLTDLDDRTTKGVSLGGAARTLN--SEFEASLNGIE--RTL 435
OY 125 AE-----YETKMAADLMAALQDMERLANSNPNHTEEVNNIKKALEAKD 169
DB 436 AERGALISEFQTRAEDALDTGOKLNAALEARARQIN---ETLVERAREIAHTFAESKD 491
OY 170 TIDKLNKLVTLQONKSLTEVLTQDSADQIPAINSOLE-----INKNSA 214
DB 492 T-----LAAMIDQCK--TQT--GADMADIVTSSMLEFARSDFRAGKEAARHYVSSSF 541
OY 215 DOIIRDLERONISYEAULTNAGEVIKASEGIRLGOALQSIYDAGDQSOAAVLAQAOQNN 274
DB 542 DSDIQLRLADARQIEAEVNHRSKRLSESD-----RMAAAMQADL-- 581
OY 275 SPDNIATKELIDAAETKYNELKQHTGLTDSPLYKKAEEQISOAKDIOETKPSGSDIP 334
DB 582 --EKFAESRDGIDAAVT--NQVQSLAEGRS--LIARALEE-----DLRKVNESRAID 628
OY 335 IYVPSGSAASAGSAGALKSSNNSGRISLLDDVQVNEAALQGRFMSIEQF--NYNN 391
DB 629 -----ASLQSHLERLEEGRN--RLSLAL--NEDSGLIVQA-RTTIDEMVAGHVAK 673

Qy	392	PATAKEL--QAMEQOLPAMSQVLVAGGAEIPAEIQAKD---ALAOALKPASDGLATAM	446
Db	674	LAEGRNLSRALEDLCKLSDSRASIDGLVAGQVEKTAEBRAVATKLE-----	722
Qy	447	GOVAFAPAAKVGSGSAGTAGTVQNMVKOLKYTAFFSTSSSYAAALSPGVSAYVTKLSLVS	506
Db	723	-----NDIAGIKSLIEVHSAKLTVE-----DRQLGRSLREDDLSGINGLLDHS	765
Qy	507	ESRSGVQASISQOTANPALSR--SVSRSGIE-----SQGSADASQAAET-----	549
Db	766	VKLANDRSLLSQTEADLAKTAEBRSRSDIDGLVAGQVEKLEAGR--DILKALRSLSDLTIK	823
Qy	550	--IVRDSQTEFLDVSRTL--QVLDLSMTSTIVSPQANQOEIMQKLTASIKSPQFGYPAYON	606
Db	824	GVISQSGSRLEADNEDRGQLSRVLEADLQWNVNSVIADHNRRLYVD--RSLTKSLLEDLDAKLMD	882
Qy	607	SADSLQAKRAQLEPEVDG---ERSLAESQENAFKQPAFIDQVLYVNIASLFS	656
Db	883	SRSSIDGLVAGQVEKLEAGRDILKRAL--EADPLNTIKTVAQDSQKLLDDDAQCA	935

RESULT	38		
049526			
ID	049526	PRELIMINARY:	PRT: 716 AA.
AC	049526;		
DT	01-NOV-1996	(TREMBLrel. 01, Created)	
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)	
DT	01-NOV-1998	(TREMBLrel. 08, Last annotation update)	
DE	lmp1.		
GN	lmp1.		
OS	Mycoplasma hominis.		
OC	Bacteria; Filicutes; Bacillus/Clostridium group; Mollicutes;		
OC	Mycoplasmataceae; Mycoplasma.		
OX	NCBI_TaxID=2098;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MH81;		
RC	MEDLINE=95369883; Pubmed=7543881;		
RA	Jensen L.T., Ladefoged S., Birkeland S., Christiansen G.,		
RT	"Selection of Mycoplasma hominis pc21 deletion mutants by cultivation		
RL	in the presence of monoclonal antibody 552."		
RL	Infect. Immun. 63:3336-3347(1995).		
SE	EMBL; U21963; AAA81014.1;		
SO	SEQUENCE 716 AA; 79864 MW; 28D08C3E0C91CB47 CRC64;		

Query Match 6.7%; Score 214; DB 2; Length 716;
Best Local Similarity 20.9%; Pred. No. 0.36;
Matches 128; Conservative 122; Mismatches 241; Indels 120; Gaps 24;

Oy	61	SANALISLRA-----ILNKSNSPPDLSLOLEASSTYSTVYRVAKVDDEAKSNPDA	114
Dd	162	SLESLLELTIONATNNILKLENESTREKONIDNLAK-----EOLKASISOA	207
Oy	115	KSGLENAKTTAEYETKMAIDMAALQDMERLANDSPDSNNHTTEEVYNNIKKALEAKODTIDKL	174
Dd	208	NOLLPO---LSDNDSEIKAKKSLDAETIKANANOVAASNNTASMQAKSSLDKAEIITK-	263
Oy	175	NKILYLOQNSLLEVEYKTTSDAQIPAINSOEINKNSMD--QIYKDERONISYEAVL	232
Dd	264	KLETFNKDKAKFNEIKOTRN-----QIOEFITNNNNRYSLSLSQLSKSDKSNSTV	317
Oy	233	--TNAGEYIKASSEAGIKLGOALQSIYDAGDOSAAVLOAOONNSPDI--AATKELIDAA	289
Dd	318	DSMNSDIESANTELKQALAKANDKYOADBLAKS--IKELNNSVSANATLTSAKLTD--	373
Oy	290	ETKVELKOEHTGLDPSLYVAKAEEOISOAKODIOELKIPGSDPIYGPBGSAASGSA	349
Dd	374	--KDNITQAAETL-----EKEDYQAOAIK-----SNNRASHMQSAK	408
Oy	350	GALXSSNNSGRISSLDDVDNEMAIALQGEFSMIEOF--NWNPNPATEKLOAMEAOULT	406
Dd	409	SSLAKVAEIRKKLETFNKDEAKFNEIKOTRNOIOEIFITNNKNP-----NYSGLLSOULT	464

QY	407	AMSQJLVG--ABGEIPEAEIOAKIKDLAQAALQPSADJGLATMGQVAPPAAKVGGSGSACTAG	465
Db	465	SKROSKNSVTSDSNKSDIESANTELKQALAKKANADQQA--DNLASIKEQJLNNVSNAN	522
QY	466	IVQMVN-----KOLYK-----TAFSTSSSSYAALSDQYSAYKTU	501
Db	523	TLSAKLTJDKDNTIQQAKTLEKEEVQKQAQNAIKSNNTASMQSAKSSJDAKVAEITKLETF	582
QY	502	N-----SLYSBSRGVQASISOTA-NPALSRKSVSNKSGIESOGRSDASQRAEETIYRD	553
Db	583	NKDKAEKFNELKQTRNQOEETINNKNKNPNYSELISOL-----TSKROSKNSVD	632
QY	554	SOTLGDVYS-RLOYLSDLSMTIYSVSNPOAQ--EEIMOKLTASISKAPQFQYPAVQNSADS	610
Db	633	SSNKSNDIESANTELKQALAKKANADKQVADNLAKSIKREQJLNNVSNANTLS-AKLTJDKDNT	691
QY	611	LQKFAAQLERE	621
Db	692	IQQAKTLEKE	702

```

RESULT 39
O9CF64
ID O9CF64 PRELIMINARY; PRT: 1072 AA.
AC O9CF64;
DT 01-JUN-2001 (TREMBLrel, 17, Created)
DT 01-JUN-2001 (TREMBLrel, 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Unknown protein.
GN Y06 OR L1617.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=114403;
RX MEDLINE=21235186; PubMed=11337471;
RA Boletín A., Wincker P., Manger S., Jallón O., Malarne K.,
RA Weissenbach J., Ehlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403."
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006392; AAK05715.1;
DR PRINTS: PR01608; BACINVAS1NC.
KW Complete proteome.
SQ SEQUENCE 1072 AA; 113056 MW; 464446E2656CA08 CRC64;

```

Query Match	6.7%;	Score 214;	DB 16;	Length 1072;
Best Local Similarity	19.7%;	Pred. No. 0.59;		
Matches 154;	Conservative 137;	Mismatches 280;	Indels 212;	Gaps 31;

QY	20	SSMNPJINGQJANSFETREKRAEABSPASSSVSSMSPFLSSAKNALISRLDILKKNSS	79
Db	295	SSSSNVEQVDVAASSESTODANASALRYPISEASSVT-----DFTLSSIS	338
QY	80	PTDSL\$OLEASTSTSTYRYVAKDYDEAKSNFDPKAGLENAKTLAEYTKMADLMAALO	139
Db	339	LDSSIS\$SQTEHNSOGASSTAEISYDSEBNS- SLSSNOJNNSNSENKEDNO\$SLSSMS	397
QY	140	DMERLANDPSN-----NITEEVNMI-----KKALEAKQDITDKL-----	174
Db	398	SNESEHNSN\$NINETNNSSEITNLPSPNPTESNVSDDTSSPASTN\$N\$ISLSP\$NI	457
QY	175	-----KLVLTQ\$ONK\$SL\$EYVLTQD\$ADQJPAIN\$OLEINK	211
Db	458	STSDSE\$ATNSDPSNVAEYVANS\$LASVNNSS\$-----VLS\$T\$TAD-----NLGINO	507
QY	212	NSADQITND-----LERONISLEAVLTNAGEVYKAS\$ENGIRKLGALOSYDADG	261
Db	508	SG\$DMLTUD\$SEIST\$GAF\$LS\$NOT\$S\$EA- \$T\$N\$N\$S\$SI- \$SL\$PS\$N\$IS\$T\$VLE\$T\$T\$SS\$N	565
QY	262	Q\$OAAVILQ-----AOONNSP\$NIAATKELIDIAAE\$K\$VNE-----LKOEHTGLT\$P\$PLVKA	312

```

Db      566 FSNVAEVANNSLASVNNSSSVLSSTSTADNLE--INQGSCLNLTKRDSEISTSGAFLSS 623
Qy      313 EEOISOAKD-IOETKPSGSDPIPIVPGSASAGSAGALKSSNSG-RISLLDDVDN 370
Db      624 NOTSEASNSMSINSPLSLTSLNSSEATQNSSEATQVNNSSSHSSNILLNSGN 683
Qy      371 EMAIALLOGFRSMIEQFNVNPNPAKEIQAMEAQLTAMSDQLVGADGELPAETQAIKDAL 430
Db      684 DSSD---SDSDSDSDSSNLSPPNLETNQTISSK-----PSEVNNTISENP 725
Qy      431 AALALOPASDGLATAMGOVAFAAVGGSGAGTACTVGM-NYKQLYKT-----AFSS 481
Db      726 KVVSSSNVQENSTHENSSTPNKSSITSPITSTSSQKESQSNLNTTEGINNPTFFNN 785
Qy      482 TSSSSYAALSDGYSAKYTLNS---LY--SESRGVOGAIQOTANPALRSRVSRSIGIS 535
Db      786 SSEENSAASILTSYGNNSSESEFCTCLYSNENQDNGSEISHSLPSSNSNNVSSIOS 845
Qy      536 Q-----GRSADASQRAAETIV-----RDSQT-----LGDVYS----- 562
Db      846 QALLESSKSTYKRRSSSLIISTINSHQMEDNQNSNDEKSNNNVESITLQNLNISTNKT 905
Qy      563 -----RLQVLDLSMS-----TIVSNPQANQ-EIMQKTRA---S 592
Db      906 HKNSTLSQKLSVITLTPSKSKVTNEKNENSNTVSEKILKTQKNDENSOILQITLADLS 965
Qy      593 ISKAPQFGPAYQNSADSLQKFAQLEREFEVDGERS-----LAESQENARFKQAPATIO 645
Db      966 FKK-----EYETMEDSSTVDPKVLNDE--NGDRSQNNKJTITAKDKKVFYFKRSEFNS 1016
Qy      646 QVL 648
Db      1017 KIL 1019

RESULT 40
Q8T805 PRELIMINARY; PRT; 1489 AA.
AC Q8T805;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE SD05887P.
GN CG3493.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celisier S.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY075577; AAL68382.1;
SQ SEQUENCE 1489 AA; 170145 MW; 876DC2DE469945AF CRC64;

```

Query Match 6.6%; Score 213; DB 5; Length 1489;
 Best Local Similarity 19.1%; Pred. NO. 0.98;
 Matches 141; Conservative 127; Mismatches 288; Indels 184; Gaps 23;

```

Qy      5 HHMHESGPESVSNQSNMPTINGQIASNSETKSTKASEASPSASSVSWSPFLSSAKN 64
Db      326 HTELEKQTEVFKLQEKLEKQLESQRESHNNEVKEQFKLQATKQEVDAK-----LMATEH 380
Qy      65 ALISLRDALINKNSPTBSLQLEASTSTSTYTRV--AAKDYDEAKSNPDYAKSGLENAK 122
Db      381 LNTLKESYAIKEQGVTLLEQLA-----IRVENEQKVDLOKONEDRNTQASDSE 433

```

```

Qy      123 TLAEYETKMAIDMALQDMERLANS----- 147
Db      434 QLKTLQAAVOADESOLLSKQDLLESIRSEQAQKQOLKHLKQELKQKQENENYLDKIRE 493
Qy      148 --DPSNNHTEEVNNIKKALEAQKOTIDKLNKY-----TQONQNS-----LTEL 191
Db      494 SKSSDSQTNQADQOKTLQAAKDEAE--SKLLATEELLHSIRNDYKQAEKVALLDEKL 551
Qy      192 KTTD-----SADQIPALNSOLEINKNRSADQILKDL-----ERONISYEAVLTNAGEVI 239
Db      552 KTLSEKENVNVAKLIHINQREAOQSTDSQKTNELRAKDEAFALISTHSLNLQAL 611
Qy      240 KASSEAGIKLQAL-----QSTYDA--GDQSOAAVLQAOQNNSPDNIAATELIDA 288
Db      612 SAKERQASLEQSLNAKLTSESHSLQDLRLHNDQLEIYQRRHQNDWEAQLARAREELAA 671
Qy      289 AET-----KYNELKQEHGTGLDSPLYKABE--QISOAKQDIQKIPSSDPI 335
Db      672 IQSREHLALELEKSLMEKRESVAALNSEKASQEOHRLKLEQILQREILOIDQDHAN-- 728
Qy      336 VQPSGSAASAGSAGALKSSNNSGRISLLDDVDNEMAAIALQFRSMTEQFPVNNPATA 395
Db      729 -SESEYVALKQQLDALSDQLATSQASILLAKKEKELKASGNKLNKIKQHOHQAKSSDS 787
Qy      396 KELQAMENQLT--AMSDQLVGADGELPAETQAIKDALAQALKQPSADGLATAMGOVAF 452
Db      788 VLELALQSLADRLSHSRQVESEKEELQARVGIIEI----- 825
Qy      453 AAKVGGSGAGTQGMNKKLYKTAFSTSSSSYAALSDGYSAKYTLNLSYSE----- 507
Db      826 -----GTMOAQMOQV-----ODSHSELERERKKLESRESLQOQVDS 864
Qy      508 ---SRGVOAISQOTANPALSR-----SVSRSGIESQGRSADASORAAETIVRDSQTLGD 559
Db      865 AADERTSAKLEIEIOSENTLAEERNCNLEBQANHLESQ--LOAKQDEIGIKAKIQOVLDD 921
Qy      560 VYSRLQVLDLSMS---TIVSNPQA--NOEIMQKLTASTSKAPQFGYPAYQNSADSLQK 614
Db      922 EHSKIQNQLMDHDHRTLODKDAYEKDKLTKHTLD-----CLQASASELHRV 971
Qy      615 AQLEREFPDGRSLAESQ 634
Db      972 KANLDRELKEQDQOLSEURE 991

```

Search completed: March 12, 2003, 12:21:40
 Job time : 133 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 09:42:47 ; Search time 21 Seconds
(without alignments)
1303.542 Million cell updates/sec

Title: US-10-007-693-139
Perfect score: 3204
Sequence: 1 MHNNHMHESGPESVSNOS.....PAFIQOVLVNIASLPSGYLS 660

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224.5	7.0	1790	US01_YEAST	P23586 saccharomyc
2	209	6.5	1433	REST_CHICK	O42184 gallus gall
3	208	6.5	1306	MSB2_YEAST	P32334 saccharomyc
4	202.5	6.3	1966	MYSB_CAEEL	P02566 caenorhabdi
5	201	6.3	2116	MYSA_DICDI	P08799 dictyosteli
6	200	6.2	1969	MYSA_CAEEL	P12844 caenorhabdi
7	199.5	6.2	1140	YMS6_YEAST	Q04893 saccharomyc
8	199.5	6.2	1381	YBE7_YEAST	P34216 saccharomyc
9	199.5	6.2	1957	YDB6_SCHPO	Q10411 schizosacch
10	199	6.2	1509	MYSN_ACACA	P05659 acanthamoeb
11	199	6.2	1940	MYH3_RAT	P12847 rattus norv
12	199	6.2	2492	TALA_DICDI	P54633 dictyosteli
13	198.5	6.2	539	MYH3_HYDAT	P39922 hydra atten
14	197	6.1	1938	MYS_AEOIR	P24733 aequipecten
15	194	6.1	1093	TMF1_HUMAN	P82094 homo sapien
16	194	6.1	1940	MYH3_HUMAN	P11055 homo sapien
17	194	6.1	1962	MYSA_DROME	P05661 drosophila
18	193	6.0	1938	MYHD_HUMAN	Q9ukx3 homo sapien
19	192.5	6.0	1727	ALM1_SCHPO	Q9ukx3 schizosacch
20	192	6.0	1938	MYH4_RABIT	Q28641 oryctolagus
21	192	6.0	2022	ANT1_ONCYO	P21249 onchocerca
22	190.5	5.9	1462	NKCR_HUMAN	P30414 homo sapien
23	190	5.9	1939	MYH1_HUMAN	P12882 homo sapien
24	189.5	5.9	1938	MYSD_CAEEL	P02567 caenorhabdi
25	188.5	5.9	3210	CENF_HUMAN	P49454 homo sapien
26	188.5	5.9	918	YMOB_CAEEL	P49454 caenorhabdi
27	188.5	5.9	995	YIO9_YEAST	P40442 saccharomyc
28	188.5	5.9	2017	MYSN_DROME	O99323 drosophila
29	188	5.9	1528	SPAA_STRDO	P21979 streptococc
30	187.5	5.9	1935	MYSS_CYPCA	O99339 cyprinus ca
31	187	5.8	1084	MYSS_RABIT	P02564 oryctolagus
32	187	5.8	1935	MYH7_PIG	P79293 sus scrofa
33	186	5.8	697	MEP1_LYCCE	P93203 lycopersico

ALIGNMENTS

RESULT 1	ID	US01_YEAST	STANDARD:	PRT: 1790 AA.	
AC	P23586;	01-MAY-1992 (Rel. 22, Created)			
DT	01-MAY-1992 (Rel. 22, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Intracellular protein transport protein US01.				
GN	US01 OR INT1 OR YDL058W.				
OS	Saccharomyces cerevisiae (Baker's yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-X2180-1A;				
RX	MEDLINE=91185402; PubMed=2010462;				
RA	Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,				
RA	Yamasaki M.;				
RT	"A cytoskeleton-related gene, us01, is required for intracellular				
RT	protein transport in Saccharomyces cerevisiae.";				
RL	J. Cell Biol. 113:245-260(1991).				
RN	[2]				
RP	SEQUENCE OF 782-1790 FROM N.A.				
RA	Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,				
RA	Kendrick K.E.;				
RL	Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1-8 FROM N.A.				
RA	Bai Y., Symington L.S.;				
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI				
CC	COMPLEX.				
CC	- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR				
CC	MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE				
CC	ER AND THE GOLGI COMPLEX.				
CC	- DOMAIN: THE RODLIKE TAIL. SEQUENCE IS HIGHLY REPETITIVE, COMPOSED				
CC	OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL				
CC	COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.				
CC	- SIMILARITY: BELONGS TO THE VDP/US01/YBL047C FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: X54378; CA38253.1; -				P35749 homo sapien
DR	EMBL: L03188; AAB00143.1; -				Q28628 oryctolagus
DR	EMBL: U53668; AAB66559.1; -				Q10970 mycobacteri
DR	PIR: A38455; A38455.				P30622 homo sapien
DR	SGD: S0002216; US01.				P12845 caenorhabdi
DR	InterPro: IPR002017; Spectrin.				P47033 saccharomyc
DR	Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.				O99623 homo sapien
DR					P35580 homo sapien
DR					O92351 schizosacch
DR					P23504 streptococc
DR					P02564 rattus norv
DR					O27991 bos taurus

```

FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 847 847 G-> E (IN REF. 2).
FT CONFLICT 924 924 E-> K (IN REF. 2).
FT CONFLICT 1253 1253 V-> I (IN REF. 2).
FT CONFLICT 1319 1319 I-> V (IN REF. 2).
FT CONFLICT 1461 1461 N-> S (IN REF. 2).
FT CONFLICT 1581 1581 G-> S (IN REF. 2).
FT CONFLICT 1600 1600 I-> V (IN REF. 2).
FT CONFLICT 1661 1661 R-> S (IN REF. 2).
FT CONFLICT 1772 1772 D-> DEEDDEE (IN REF. 2).
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

```

Query Match Best Local Similarity 7.0%; Score 224.5; DB 1; Length 1790; Pred. No. 0.037; Mismatches 155; Conservative 130; Matches 268; Indels 189; Gaps 32;

```

QY 3 HHHHHHSGPESEY-----SSNOSMN-----PIINGQIASNSE--- 35
DB 969 NKKKMOAENESLITKAVEESKNESSIQLSNLQNKIDMSOEKFNQIERGSIKNEIQLK 1028
QY 36 ---TKESTKASEPSPASSSVSSWFLSSAKNAL--ISLRDAIINKNSPTDSLQLE 88
DB 1029 KTIISDLEQKREIISKSDSKDEYESQILKKELETAATANDENVNKISELTKREBLE 1088
QY 89 ASIST-----SYTPRAAKYIDAKSNFTPAK--SGLENATLAEYETKADLMAAIO 139
DB 1089 AELAAAKNKNLELETKETSEKALKVEKEHEHLEKEKIQLEKAEYETKQOLNSLANLE 1148
QY 140 DMRLANSPDNNHTEEVNNIKALE--AOK-----DTIDKLN-KLVTLONOKSLTEVL 191
DB 1149 SLEK-----EHEDLAOLKYEEOIANKERQVNEISQLDETTSTQOEWESIKR-- 1198
QY 192 KTTSDAQDIPAINIS---OLEINKNSADQI--TKDLERONISYEAVLTNAGEVIRASSE 244
DB 1199 KNDLEGEVAKMSTSEEGSNLKESEIDALNLIQIKELKKNETNEASLT--ESIKSVES 1255
QY 245 AGRIKIQ-----ALQSYDAGQSOAAVLQAOQNNSPDNIAATKRLIDAAET 291
DB 1256 EYVKKIQLDECNFKKEVSELEDKLKASDKSKYLELQKESK-----KIKELDAKTP 1310
QY 292 KVNELKOEHTGL-----TDSPLVKKAEQISOAKQIDQIEFKSGSDIPI 335
DB 1311 ELKIQLEKTNLSEKAKESSELSRLKTSSEERKNAEQLKKNELQ--IKNQAE--- 1366
QY 336 VGPSSGASASAGALKSSNNSGRISLLDDV---DNEMAAIALQGFRRMIEQFNVN 390
DB 1367 --KERKLLNCGSST--ITQYSEKINTTLEDELIRLQNNENELKAKEIDNTRSELEKVSLS 1421
QY 391 NPA-----TAKEIQ--AMEAQITAMSDQVLGADGELPATIQAKDLAALQPSA 439
DB 1422 NDELEEKQNTISLQDELISYKDKITRNDKELTIERDKRRLQLESEKELRAQE--- 1477
QY 440 DGLATAMQVAAFAAKAGVGAGTACTAGVQNNVQKLYKTAASTSSSSYAALISDGYSAKY 499
DB 1478 -----SKAYEEB-----LKKL-----EEESSEKREKLEKSEMMK 1508
QY 500 TLNLSYSESRSGVQSAISQITANPALSRVSRSISGSRASASQRAETIIVDSQ--TLG 558
DB 1509 KLSSTESNETELKS-----SMETIRKSDEKLEQSKSKSAEEDIKNLQHEKS 1554
QY 559 DVSURL-----QVLDLSMS--TIYSNPOANOELIMOKILITASIPQFGYAVONSADSLQ 612
DB 1555 DLISRINESEKIDLEELKSKLRIRIAKSGSELETVKQELNNAQEK---IRINAENTV--LK 1609
QY 613 KFAAOLEREFVDEGRSLAESQ 634
DB 1610 SKLEDIRLEKDKQAEIKSNQ 1631

```

RESULT 2

ID REST_CHICK

AC 042184; 042228; 057563; 057564; PRT; 1433 AA.

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE 15-JUN-2002 (Rel. 41, Last annotation update)

GN Restin (Cytoplasmic linker protein-170) (CLIP-170).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

ON NCBI_TaxID=90311;

RX [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=98137792; PubMed=9469933;

RA Gripapic L., Volosky J.M., Keller T.C. III;

RT "Cloning and expression of chicken CLIP-170 and restin isoforms.";

RL Gene 206:195-208(1998).

RN [2] SEQUENCE OF 17-1139 FROM N.A. (ISOFORMS 3 AND 4).

RP TISSUE=Rectoralis muscle; DB 1.

RA Gripapic L., Keller T.C. III;

RT "Identification and expression of two novel CLIP-170/Restin isoforms expressed predominantly in muscle.";

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

CC - FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES (BY SIMILARITY).

CC - SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE CYTOSKELETON (BY SIMILARITY).

CC - CYTOSKELETON (BY SIMILARITY). At least 4 isoforms; 1 (shown here), 2, 3/CLIP-170(11) and 4/CLIP-170(11+35); are produced by alternative splicing.

CC - SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).

CC EMBL; AF014012; AAC60344.1; -

CC DR EMBL; AF020764; AAC60345.1; -

CC DR EMBL; AF045650; AAC03547.1; -

CC EMBL; AF045651; AAC03548.1; -

CC DR InterPro; IPR000938; CAP-GLY.

CC DR InterPro; IPR001878; ZnF_CCHC.

CC DR Pfam; PF01302; CAP_GLY; 2.

CC DR SMART; SM00343; ZnF_C2HC; 1.

CC DR PROSITE; PS00845; CAP_GLY 1; 2.

CC DR PROSITE; PS50245; CAP_GLY_2; 2.

CC Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.

CC FT DOMAIN 79 121 CAP-GLY 1.

CC FT DOMAIN 144 207 SER-RICH.

CC FT DOMAIN 235 277 CAP-GLY 2.

CC FT DOMAIN 305 332 SER-RICH.

CC FT DOMAIN 351 1353 COILED COIL (POTENTIAL).

CC FT DOMAIN 1414 1427 CCHC-BOX.

CC FT DOMAIN 458 492 MISSING (IN ISOFORM 2).

CC FT VARSPLIC 458 492 TOTKLEHAIKELQSLFEKTKADLRQLEEDTR ->

CC FT VARSPLIC 458 492 RKRQISEDPEN (IN ISOFORM 3).

CC FT VARSPLIC 803 803 S -> GGSSKVS (IN ISOFORM 3).

CC FT VARSPLIC 458 458 T -> RKRQISEDPERT (IN ISOFORM 4).

CC FT VARSPLIC 309 309 K -> R (IN REF. 2; AAC03547).

CC FT CONFLICT 440 440 E -> V (IN REF. 2; AAC03548).

CC FT CONFLICT 440 440

SQ SEQUENCE 1433 AA; 161026 MW; 5631CE8683498E23 CRC64;

Query Match Best Local Similarity 6.5%; Score 209; DB 1; Length 1433; Pred. No. 0.12; Mismatches 154; Conservative 126; Matches 260; Indels 342; Gaps 28;


```

Db 302 MSAAGIASSSTDNFTMSEFTSLSTEVDSVSTVALLSAPFLQ--TSTNSFSIV 359
OY 391 NPATAKELOAMEAQLTAMSDQLVAGDELPAEIOAIKDALAOLKOPSDGLATAMGOYA 450
Db 360 SPVSVPVPSQSSSDVASSSTANV-----VSSFSDFPPTSTGCVSVYAOQA 407
OY 451 FAAA-----KVGGSAGTAGVQVMNKLYKTAFFSTSSSSVYAAALSD---GYSAVKTL 501
Db 408 SALAFQSSSTEVYGCASASTMSSTLSLTSTLQSTTLDDSSSLASSASSSDLTLDYGVSTASI 467
OY 502 NLIYSRSRGVQSAISQTNPALRSVSRGIESGCRSADAGRAAEITVRDSQTIGDYV 561
Db 468 PLLSASEQDSTSSSFS-VVSPVSEVPSQS-----SSDVASTSPVSSSFS---Y 515
OY 562 SRLQVLDSTLM-----STIVSNFOANEEIMOKLTASISRAPOFGYPAYONSADSLQFAA 616
Db 516 TSLQAGGSSMTNPSSSTIYSSSTGSSSESAASTASATL-----SGSSSTYMG 564
OY 617 QLERFVGDGERLSAESQ 633
Db 565 NLOSQPPSTSLISESQ 581

RESULT 4
MSB_CAEEL
ID MYSB_CAEEL STANDARD; PRT; 1966 AA.
AC P02566;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain B (MHC B).
GN UNC-54 OR MYO-4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83273600; PubMed=6576334;
RA Karn J., Brenner S., Barnett L.;
RT "Protein structural domains in the Caenorhabditis elegans unc-54
RT myosin heavy chain gene are not separated by introns."
RT Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
RN [2]
RP SEQUENCE OF 850-1966 FROM N.A.
RX MEDLINE=82272395; PubMed=7202124;
RA McLauchlan A.D., Karn J.;
RT "Periodic charge distributions in the myosin rod amino acid sequence
RT match cross-bridge spacings in muscle."
RT Nature 299:226-231(1982).
RN [3]
RP SEQUENCE OF 1876-1966 FROM N.A.
RX MEDLINE=8332892; PubMed=6571695;
RA Mills N., Gesteland R.F., Karn J., Barnett L., Bolten S.,
RA Waterston R.H.;
RT "The genes sup-7 x and sup-5 III of C. elegans suppress amber
RT nonsense mutations via altered transfer RNA."
RT Cell 33:575-583(1983).
RN [4]
RP FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN

```

```

CC C.ELEGANS.
CC -1- MISCELLANEOUS: MHC A AND MHC B ARE FOUND EXCLUSIVELY IN THE BODY
CC WALL MUSCLE. THEY CO-ASSEMBLE INTO BODY WALL THICK FILAMENT.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL: J01050; AAA28124.1;
CC EMBL: V01494; CAA24738.1;
CC PIR: A02992; MMYN.
CC HSP: P08799; 1MND.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00063; myosin_head; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC PRODOM: PD000355; myosin_head; 1.
CC SMART: SM00242; MYSC; 1.
CC Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
CC ATP-binding; Methylation; Alkylation; Multigene family.
CC MYOSIN HEAD-LIKE
CC DOMAIN 1 850
CC FT DOMAIN 851 1966 COILED COIL (POTENTIAL).
CC FT DOMAIN 851 1164 ALPHA-HELICAL TAILPIECE (S2).
CC FT DOMAIN 1165 1176 HINGE.
CC FT DOMAIN 1165 1966 LIGHT MEROMYOSIN (LM).
CC FT NP_BIND 177 184 ATP (BY SIMILARITY).
CC FT DOMAIN 665 687 ACTIN-BINDING.
CC FT DOMAIN 769 783 ACTIN-BINDING.
CC FT MOD_RES 128 128 METHYLATION (TRI-) (POTENTIAL).
CC FT MOD_RES 705 705 ALKYLATION (SH-1).
CC FT MOD_RES 715 715 ALKYLATION (SH-2).
CC FT CONFLICT 1337 1337 E -> L (IN REF. 2).
CC FT CONFLICT 1880 1880 I -> R (IN REF. 2).
CC SQ SEQUENCE 1966 AA; 225125 MW; B6f0BB2FE27B67F CRC64;

Query Match 6.3%; Score 202.5; DB 1; Length 1966;
Best Local Similarity 21.9%; Pred. No. 0.33;
Matches 136; Conservative 112; Mismatches 244; Indels 129; Gaps 27;

OY 82 DLSIQLEASTSTSTYRVAKYDEKSNFDTAK-----SGLENAKT-LAEYETKADLMA 136
Db 866 DKVKALEDSLAKEEKLK---KELESSAKLVBEKTSLFTNLESTKTQLSDAEERLAKTEA 922
OY 137 ALQDM-----ERLANSPPSNHTEEVNNIKKALEAQDTIDKLNKLVTQONKRSUT 188
Db 923 QOKDASKQSELDQDLADNE---DRTADVQRAKKKEALEVEALKR-----QODELSMR 974
OY 189 EYLKTTDSAD-QIPAINSOLEINKNSADQIITKLERONISYEAVLTNAGEVITAKASSEAG- 246
Db 975 KAEEKOSKDHQIRSLQDDMM-----QODEALKLNKEKKHOBEINKLMEDQLQSEEDKGN 1030
OY 247 -----IKIGQALQSLVADGDQSQ--AAVLQAQONNSPNIATKELIDIAETK----VN 294
Db 1031 HONKYVAKLEQTLDDLESLERKRRADLDQKRKVEGELKTAQENIDESQRHDLN 1090
OY 295 ELKQEHTEG-----TSPYVKKAKEPQISOAKQIOEIKSGSDIPYVGSASAAG 346
Db 1091 NLKKKSELSHVSRLDEQALYSKLQROIKDQGSRTISELEE-----LENERQSRKAD 1145
OY 347 SAAGALKSSNNGRSLILDDVDNEMAAITALQ-----GFRSMTEQFNVNP-- 392
Db 1146 RAKSDQD-----ELEELGEKLTDEGGGATAQVEVKKREAEELAKRRRLDEEANNHMQ 1200
OY 393 --ATAKELOAMEAQLTAMSDQLVAGDELPAE-IOAIKQA--LAQALKOPSDGLATAMG 447

```



```

Db 1201 LGLKKHTDAVALTLDLDQLNKAKAYEKDKAQRDAEDLAAQLDQETSGKL----- 1255
OY 448 QVAFAAKVGSGAGTAVQNNKOLYKTAFSSSTSSSSAAALSDGYSAVYKTLNLSYSE 507
Db 1256 -----NNEKLAKOFELQLTEL-----QSKADQSQQLQDFTSLKRLHSENGD 1298
OY 508 SRSQVQASISQTPALNSRSHRSRSGIESQGRSADASORAAETIYVDSQTLGIVSRLOYL 567
Db 1299 LVROLEDAESQVNLQTLRLKSQLTSLQLEARRTRADEARENRQVVAQAQKNYCHEAQLO-- 1356
OY 568 DSLMTIYSNPQANDETMOKITASIKRAPHQPRPAVQNSADLSQFPAALEREYVDGER 627
Db 1357 ESLEBEI-----EGKNELRLQL-----SKA-----NAD-IQWKARFE-----GEG 1391
OY 628 SL-AESQENAFKQAPFIQOV 647
Db 1392 LKKADELDAKRRQAKINEL 1412

RESULT 5
MYS2_DICDI
ID MYS2_DICDI STANDARD: PRT: 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCa.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
   Dictyostelium discoideum.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vietmeier D., Schleicher M., Grabatin B., Wippler J.,
   Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
   phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RL FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
   Dictyostelium myosin heavy chain.";
RL FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutcliffe K., Holden H.M.,
   Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
   discoideum complexed with MgADP, BeFx and MgADP.ALp4-.";
RL Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
   truncated head of Dictyostelium discoideum myosin to 2.7-A
   resolution.";
RL Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;

```

```

RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-ADP-vanadate complex of the
   Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RL Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gulick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammaS, and MgADPPNP complexes
   of the Dictyostelium discoideum myosin motor domain.";
RL Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bageshaw C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization
   of Mg.2'(3')-O-(N-methylanthraniloyl) nucleotides bound to the
   Dictyostelium discoideum myosin motor domain.";
RL J. Mol. Biol. 274:394-407(1997).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
   ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -I- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
   INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
   LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
   (MLC-2).
CC -I- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
   CORTEX.
CC -I- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
   MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
   SPLIT INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
   SUBFRAGMENT (S2).
CC -I- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
   CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
   CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -I- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
   THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC -I- MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDPA ATPASE
   ACTIVITY. PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
   POSITION (688).
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 1 IQ DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
-----
DR EMBL: M14628; AAA3327.1; -.
DR PIR: A26655; A26655.
DR PIR: S00250; S00250.
DR PDB: 1MMA; 03-DEC-97.
DR PDB: 1MMD; 17-AUG-96.
DR PDB: 1MMG; 03-DEC-97.
DR PDB: 1MMN; 03-DEC-97.
DR PDB: 1MND; 17-AUG-96.
DR PDB: 1MNE; 17-AUG-96.
DR PDB: 1VOM; 23-DEC-96.
DR PDB: 1LVK; 28-JAN-98.
DR Dictydb: DD01008; mhca.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; Myosin_head.
DR Pfam: PF00063; myosin_head. 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD00035; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.

```


FT DOMAIN 857 1969 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP.
 FT DOMAIN 667 689 ACTIN-BINDING.
 FT MOD_RES 770 784 ACTIN-BINDING.
 FT MOD_RES 130 130 METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 707 707 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 717 717 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1969 AA; 225509 MW; 64577BBAF7EAD80A CRC64;

Query Match 6.2%; Score 200; DB 1; Length 1969;
 Best Local Similarity 18.9%; Pred. No. 0.42;
 Matches 142; Conservative 152; Mismatches 304; Indels 154; Gaps 28;

QY 28 GGIASSEKREKSTKASEAPSSSSVSSW-SFLSSAKNAL-----ISLRALILNKNSPTDS 83
 DB 863 GELAVKIQLE--EAVQGEIARISQLESOVADIVEEKALFLSLETERKNLADAEERNK 920
 QY 84 LSOLEASTSTVTYRVAADYDEAKSNFDTAKSGLENATLAETKMDLMAALQDMER 143
 DB 921 LNLQKA-TLESKLSDTGTGLEDMQERNEDLARQKKTDELSDTKKHVQDDELSLRKAKQ 979
 QY 144 -----LANDPS-----NNTEEVN-NIKKALEAQDTIDKLNKL--- 177
 DB 980 EKQSRDHNIRSLQDEKANDQEVAVAKLNKEKKHQEESNRKLNEDLQSEEDKVHLEKIRNK 1039
 QY 178 -----VTLQONKSLTEVLK-----TTSDAQDIPA----- 202
 DB 1040 LEOQDELEENIDREKRSKGLIEKAKRKVEGDLKVAQENIDETTKOKHDVETTLRKKEED 1099
 QY 203 ---INSOLEIRKN---SADQIIRKLERONISYEAVL---TNAGEYIKASSEAGIKLQQA 252
 DB 1100 LHHTNAKLAENNSITAKLQRLKELTARNABELEEELEARNRSKOKSDRSSEERLEEL 1159
 QY 253 LOSYDAGQSOAAVYLOAQONNSPDNIATKELIDAA---ETKYNELKOEH---TGLTD 305
 DB 1160 TRLRDOQG-ATAQOLEAKKRKEAEIYAKLRREKEEDSLNHETALISLRKHGDSVALTE 1218
 QY 306 -----SPLVKRAEEOISQOKDIOE-----IKRSGSDIPVGPS 339
 DB 1219 QLETLQKLAKKSEAEKSKLQRLDEESQAHATDSEVRSQRLKALKTIEVOYSELQTKADE 1278
 QY 340 GSAASGSAAGALKSSNNSGRISSLDDVDNEMAAI--ALQGRSMIEQFNVN----- 391
 DB 1279 GSRQLODFALAKNLNENNSDLNRSLEMDQNLNLRKSTLQSLDPTRRNYDESRK 1338
 QY 392 ----PATKELQAMEAQLTAMSDQLVGADGELPAEIQAIKDALQALQKPSADGLATANG 447
 DB 1339 RQALATATKLNLEHENTILREHLDEEAESKADLTROIISKLNAEIQOKKARFDSGL-NKLE 1397
 QY 448 QVAFPAARVGGSGAGTAGVQNMVKOLYKTAFFSTSSSYAAALSDGYSAKYTLNLSYSE 507
 DB 1398 EIE-AAKK-----ALQKLVQEL-----TDINEGIFAKIAISOEKRFKLMODL-DD 1440
 QY 508 SRSGVQSAISQTPANPALSRVSRSRGIESQGRSADASQRAETIIVRDSQTLG-DVYSRLQV 566
 DB 1441 AOSDVEKAAQVAFYEKHRQFESIIEWKKTDLSSLELDAQNRNQLDPTLFRAKTA 1500
 QY 567 LDSLSTIYSNPQAND--EIMOKLTASISKAPQEGYPAVQNSADSLQFAAQOLEFEVD 624
 DB 1501 NDELLEYLDSTRRKNSLQAEVKDLTDQEGEGR-----SVAELOKIVKLEVEKEE 1552
 QY 625 GERSLAESQENAFKROPAPFIOOVLVNIALSIF 656
 DB 1553 LOKALDEA-EALAEAEAKVLAQIEVSQIRS 1583

RESULT 7
 YM96_YEAST
 ID YM96_YEAST STANDARD; PRT; 1140 AA.
 AC 004893;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE Hypothetical 113.1 kDa protein in PRE5-PET4 intergenic region.
 GN YMR317W OR YM9924.09.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Churcher C.M., Louis E.J., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC -I- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL; 254141; CA90835.1; -.
 DR SGD; S0004936; YMR317W.
 KW Hypothetical protein; Repeat.
 SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 6.2%; Score 199.5; DB 1; Length 1140;
 Best Local Similarity 19.1%; Pred. No. 0.23;
 Matches 124; Conservative 109; Mismatches 278; Indels 139; Gaps 17;

QY 12 PPSVSSNOSSMNP-----INGQIASNSEKREKSTKASEAPSSSSVSSMSFSL 60
 DB 64 PEFITSSSLSPSTIASILSESLVIFSSLTSSDISSTVNDVSSSTSGSNSYSALS 123
 QY 61 SAKNALISLRDALILNKNSPTDLSQLEASTS--TSTVTRVAARDYDEAKSNF--DPAK 115
 DB 124 SP-----NAQLSSSTETDTSISSAIOFSPTQSSNGGSSSEPLGKSVLETPAS 175
 QY 116 SLENAKTLAEYETKMDLMAALQDMERLANDPNSNHEEVNNIKKALEAQDTIDKLN 175
 DB 176 SSDTAAVTSSTPTT-----LTDVSSSPKISSGSAVTSVGTSDAKRE----- 218
 QY 176 KLVTLQONKSKLTEVLKTTDS-----ADQIPAINSOLEINKNSADQIIRKLERONISYE 229
 DB 219 ---VFSSSTSDVSSLLSTSSPASSTISSETLPSSITLSTSPVSSSEAPKATSSVSSE 275
 QY 230 AVLTNAGEYIKASSEAGIKRLQALQSIYDAGDOQAAYVLOAQONNSPDNIATKELIDAA 289
 DB 276 ASSSTSSSV---SEAPL---ATSSVSSSEAPSSSTSSVSSSEAPSSSTSSVSEISPTT 328
 QY 290 EKKV-----NELKQETIGLTDPLVYKA-----EQIISQA 319
 DB 329 SSSVSSSEAPLATSSVSSSEAPSSSTSSVSELSSTSSVSSSEAPLATSSVSSSEAPST 388
 QY 320 QKDIOEIKPS--GSDPIYGPSSAASGSAAGALKSSNNSGRISSLDDVDNEMAAIIL 377
 DB 389 SSSVSSSEAPSSSTSSVSSSEAPSSSTSSVSELSSTKSSVMSSEVSNATSLVSSSEAPPAI 448
 QY 378 QGRSMNIEQFNVNNPATAKEIQAMEAQLTAMSDQLVGADGELPAEIQAIKDALQALQK 437
 DB 449 SSLASS-RLFSSKNSTSVSTLVATEASSVTS-----LAP 482
 QY 438 SADGLAT-----AMQVAFPAARVGGSGAGTAGVQNMVKOLYKTAFFSTSSSYAA 489
 DB 483 SSETLASNIISSSLSTGNTSVSTTSAASSTLSKSSVSNMATSTSTSSSDLSKS 542
 QY 490 ALSDGYSAKYTLN-----SLYSESRGVQSAISQTPANPALSRVSRSRGIESQGRSADASQ 544
 DB 543 SVIFGNSSTVTTSPASISLITASPLPSVWSDLTSSSEASSISNLASSAPSDNNNTIASA 602
 QY 545 RAEETIVRDSQTLGQVYSHLQVLDLSMTSTIVSNPQANDQEIQKLTASIS 594
 DB 603 SLIVTKTKNS-----VSSIVSSITSETTNEINLATSTSLIS 641

RA Skelton J., Slimmons M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA Meltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Mostl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez J., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe.",
RL Nature 415:871-880(2002).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 270690; CAA94624.1; -
CC Hypothetical protein.
KW
SQ SEQUENCE 1957 AA: 222785 MW: 3F480CA06171D9DA CRC64;

Query Match

6.2%; Score 199.5; DB 1; Length 1957;

Best Local Similarity 19.0%; Pred. No. 0.44;
Matches 144; Conservative 132; Mismatches 279; Indels 201; Gaps 28;

OY 23 NPITNOIANSSTKSTKSEASPESSSSVSMSPLSSAKNMLISRDAILKNSSPTD 82
DB 153 NGILISQQLSSNKKDKNT-----SSVTTLTSEEDVSYQKRLTNESNFSKQSPAYD 205
OY 83 SLSQLASTSTVTTR---VAAKDYDEAKSNFDTAKSGLENAKTLAEYFKMDLMAALQ 139
DB 206 LSNQL-----LVTEKLDKKEKDYKIKEDVSSIKA-----SLAEQASNKSLSGEQE 253
OY 140 DMRLANSDPSNNHTEEVNNIKALEAKQDT-DKLNKLVTLQONKSLTEVLTQDSAD 198
DB 254 RLEKLIVS--SNKTVSTLROTFENSLRAECKTLOEKLEKCAINEDSDLEELKHNV---- 307
OY 199 QIPAINSOLEINKNSADQIIKDLERONISYEAVLTNAGEYTKASSENGITLGAALOSIVD 258
DB 308 ---ANYSDAIVHK---DKLIEDLSTR-ISEFDNLKSERDLSIKNE--KLEKLRNLTIG 357
OY 259 AGDOSAAVLAQAOQNN---SPDNIAATKELIDAETKVNELKQEHGTLTDS----- 306
DB 358 SLKSDSTNSQLEENVELKESNRTHSQULTD-AESKLSFEBQNKSLKSIDETQONLS 416
OY 307 ---PLYKAEEOISAO-----KDIOIKIPSGSDIPIVPGSGA 342
DB 417 SKDKMVKOVSSLEEARSLAHATGKLAETNSERDFONKRIKDFEKLEQD-----R 468
OY 343 ASGSAAGALKSSNNGRISLLDDVDNEMAAIALOGFRSMIOFVNNPATAKELQAME 402
DB 469 ACINSSSNELKEK---SALIDKRDQE---LNNLRQIKQKRVSESTSSLSQSIQ 517
OY 403 AOL-----TAMSDQLVGADGELPAEIIQIKDALQALQPSADGATLTMGVAFAPAAK 455
DB 518 ROLLNKKKKHEVESQLENEKGELOTEIS-----NSEHLSOLS--TLAAEK 562
OY 456 VGGSGAGTACTQVMNVKQLYKTAFFSSSTSSSYAAA-----LSDGYSAYTTLNLSYSESR 509
DB 563 EAAVATNNELSESKNSIQTLCAFOEKLAKSVQLEKNEQNFSDLTSSFKLKNESHQELE 622
OY 510 SGVQSAISQ---TANPALSSVSRSG-----ISEQGSADAASQR 545
DB 623 NNHOTTIKOLKDTSSKIQLOLERANFEQEKSTLSDENNDRTRKLLKLEESNKSLLIKOE 682

OY 546 AAEIVRDSOTLGD-----VYSRL-----QVILSL----- 570
DB 683 DVDSLEKNIOTLKEDLRKSEALRFSKLEKALNREYIDNKKGHETLEAQRNDJHSLSD 742
OY 571 ---MSTVSNPANOBEIMOKLTASI-----SKAPQFGYPAYONSADSLQKFAQLERE 621
DB 743 AKNTNALISELTKSEEDVRYLRLANEYTLQDSKAKKQSFSLVNSYQSISNLHYELRDD 802
OY 622 FV-----DGERSLAESOPNARKPAPFIQOV 647
DB 803 HVNMSQNTLTLESSESKLATDCEINLTQONMTLIDNV 838
RESULT 10
ID MYSN_ACACA STRAND: PRT; 1509 AA.
AC P05659;
DT 01-NOV-1988 (rel. 09, Created)
DT 01-NOV-1988 (rel. 09, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
OS Acanthamoeba castellanii (Amoeba).
OC Eukaryota; Acanthamoebidae; Acanthamoeba.
OX NCBI_TaxID=5755;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=87308395; PubMed=3040773;
RA Hammer J.A. III, Bowers B., Paterson B.M., Korn E.D.;
RT "Complete nucleotide sequence and deduced polypeptide sequence of a
RT nonmuscle myosin heavy chain gene from Acanthamoeba: evidence of a
RT hinge in the rodlike tail.";
RL J. Cell Biol. 105:913-925(1987).
CC -I- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -I- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
CC (MLC-2).
CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC OF ALPHA-HELICAL STRUCTURES. THIS REGION IS
CC INTERRUPTED BY A HINGE AND JOINED BY A NONHELICAL TAILPIECE WHERE
CC THE REGULATORY PHOSPHORYLATION SITES RESIDE.
CC -I- MISCELLANEOUS: THE HINGE REGION MAY PLAY A KEY ROLE IN MEDIATING
CC THE EFFECT OF HEAVY CHAIN PHOSPHORYLATION ON ENZYMACTIC ACTIVITY.
CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -I- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y00624; CAA68663.1; -
DR PIR: A27224; A27224.
DR HSSP: P08799; IAMD.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 2.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Colled coil; Actin-binding; ATP-binding; Calmodulin-binding;
KW Methylation; Alkylation; Phosphorylation; Multigene family.
FT DOMAIN 1 789
MYOSIN HEAD-LIKE.

FT DOMAIN 790 819 IQ.
 FT DOMAIN 848 1509 COILED COIL (POTENTIAL).
 FT DOMAIN 848 1226 ALPHA-HELICAL TAIPIECE (S2).
 FT DOMAIN 1227 1252 HINGE.
 FT DOMAIN 1253 1509 LIGHT MEMBRANOSIN (LMN).
 FT DOMAIN 1253 1482 ALPHA-HELICAL TAIPIECE (LMN).
 FT DOMAIN 1483 1509 NONHELICAL TAIPIECE.
 FT NP_BIND 182 189 ATP.
 FT DOMAIN 660 682 ACTIN-BINDING.
 FT DOMAIN 766 780 ACTIN-BINDING.
 FT MOD_RES 133 133 METHYLATION (TR-1) (POTENTIAL).
 FT MOD_RES 700 700 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 1489 1489 PHOSPHORYLATION.
 FT MOD_RES 1494 1494 PHOSPHORYLATION.
 FT MOD_RES 1499 1499 PHOSPHORYLATION.
 SQ SEQUENCE 1509 AA; 171201 MW; 2CE49BE51173D17E CRC64;

Query Match 6.2%; Score 199; DB 1; Length 1509;
 Best Local Similarity 20.4%; Pred. No. 0.33;

Matches 148; Conservative 124; Mismatches 276; Indels 176; Gaps 30;

QY 17 SNOSNMPIINGOLASSEKSESTKASPSASSSSVSFLSAKKNALISRDALINK 76
 DB 740 STDAVKADLEHLIA-----KEPTKVAK--DEVRFVTKLFFRSGLAIEELRE----- 787
 QY 77 NSSPTDLSOLEASTSTSTVTRVAADYDEAKSNFDTAKSGLENAKTLAE----- 126
 DB 788 -----QAKSMVVSIOGARAFIARRYDMRGQVYSAKILQINIRAMELKNMAMWQLY 842
 QY 127 -----YETKADLMALALQDMER-LANDSPSNHTEEVNNTIKKALEAQKDTDK- 173
 DB 843 VKARPLISQRFKEIDDLKQYKDELEKLAALKAN--AKLDKEQLAEEDADKLEKO 899
 QY 174 -----LNKLVPLQONKSLFEVLKTTDSA--DOIPAINSOLEINKNSAQDII--KDLER 223
 DB 900 LAALKLKIIDLEGKADLEE-----DNALQKRYAGLEELQETSANLILQKRLTA 954
 QY 224 QNITSYEAVALTNAGEVITKASSEAGIKLQALQSIYDAGDQSOAA--VLQAOONNSPDNIAA 281
 DB 955 EKGLKASTLEEERNRKALQEAKTQVSEERNELODKYEDAAHDSLKKEEDLSRELRE 1014
 QY 282 TKE-LIDA-----AEKVNELKQEHGTGLD-----SPLVKKAEQISOAQKD 322
 DB 1015 TKALADLANISSETLRKSLKLNTERGADVRNEIDVTATKLOLEKTKSLSEBELAQTRAQ 1074
 QY 323 IOETKPSGSDIPYGPSPGSAASAG-----SAAGALKSSNNSGRISLLDDVDNEMAAI 375
 DB 1075 LEEK-SGRE-----AASSKKQQLQOLEDARSEVDSLAKSLAAEKSLTKAKQON----- 1124
 QY 376 ALQGRSMIEQFNVNNDATA--KEIQAMEAQLTAMSDQIVGADGELPAEIQAIKDALAQ 432
 DB 1125 -----RLDRLQLEDEERTVRANVDKOKKALBAKLTLEDQYATLDGQKNA-----AAAQ 1172
 QY 433 ALKQPSADGLATAMQVAFAPAAKVGSGSAGTAGVQNNVQIKYTAASSIS--SSSTAA 490
 DB 1173 A-----KTKTYQVDETKRRLKEAEASAA-----RLKERKNALDEVAQLADDAE 1218
 QY 491 LSDQSAKYTLNLSYSESRGVOA-----ISQANPAL 524
 DB 1219 RDSAQQRKNTIRISLQLEBELAPKTGGASBEVKRLGELERLEELLTQAEARAAA 1278
 QY 525 SRSVSRSGISQ--GRSADASQRAETIVDSQTL-GDV-YSRLOV-----LDS-- 569
 DB 1279 EKNIDKANLELEELRQADDAARNDKLVKDNKRLKADLEARQOLEEEDDAKSHADSS 1338
 QY 570 --LMSTVSNPQANQEIIMKLTASISKAPQFGPVPANQNSADSIQKRAQLERFVNGER 627
 DB 1339 RRLAELEELKRYAKETSPDKQKQDKA-----NYQRENSLIKADRDSTERRNRDAER 1392
 QY 628 SLAE 631
 DB 1393 QVRD 1396

RESULT 11
 ID MYH3_RAT STANDARD; PRT: 1940 AA.
 AC P12847;
 DT 01-OCT-1989 (rel. 12, Created)
 DT 01-OCT-1989 (rel. 12, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Myosin heavy chain, fast skeletal muscle, embryonic.
 GN MYH3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87060988; PubMed=3783701;
 RA Strehler E.E., Strehler-Page M.-A., Perriard J.C., Petrasamy M.,
 RA Nadal-Ginard B.;
 RT "Complete nucleotide and encoded amino acid sequence of a mammalian
 RT myosin heavy chain gene. Evidence against intron-dependent evolution
 RT of the rod.";
 RL J. Mol. Biol. 190:291-317(1986).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEMBRANOSIN (LMN) AND 1 HEAVY MEMBRANOSIN (HM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: X04267; CA27817.1; -
 CC PIR: A24922; A24922.
 CC HSSP: P13538; 2MYS.
 CC InterPro: IPR000048; IQ_region.
 CC InterPro: IPR004009; Myosin_N.
 CC InterPro: IPR002928; Myosin_tail.
 CC InterPro: IPR002017; Spectrin.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF00612; IQ; 2.
 CC Pfam: PF01576; Myosin_tail; 1.
 CC Pfam: PF02736; Myosin_N; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC SMART: SM00015; IQ; 1.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PS50096; IQ; 1.
 CC
 CC Myosin: Muscle protein: Coiled coil: Thick filament: Actin-binding;
 CC Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 CC Multigene family.
 CC
 CC FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
 CC FT DOMAIN 782 811 IQ.
 CC FT DOMAIN 840 1933 COILED COIL (POTENTIAL).
 CC FT NP_BIND 179 186 ATP (POTENTIAL).
 CC FT DOMAIN 656 678 ACTIN-BINDING.

FT DOMAIN 758 772 ACTIN-BINDING.
 FT MOD.RES 130 130 METHYLATION (TRI-) (POTENTIAL).
 FT MOD.RES 696 696 ALKYLATION (SH-1).
 FT MOD.RES 706 706 ALKYLATION (SH-2).
 SQ SEQUENCE 1940 AA: 223857 MW: 8505465963569696 CRC64;

Query Match 6.2%; Score 199; DB 1; Length 1940;
 Best Local Similarity 21.5%; Pred. No. 0.46;
 Matches 155; Conservative 114; Mismatches 271; Indels 180; Gaps 33;

OY 8 MEGPSSVSSNOSMPT---INGQTA---SNSETKES-----TKASAPSP 48
 DB 1232 LSSSVSVSKANLEKICRTLEDQSEAKNEEFORSSELTQKSRLOTEGELISQ 1291
 OY 49 ASSSVSWPSSAKNA---LISLADALINKSSPTDLSOLEASTSTVYRAVAKOY 104
 DB 1292 LEKESIVSOLSKQAFQOIEELKROLEBEKKANNAALAHQSSRHDCDLREQEYE 1351
 OY 105 DEAKSNFDTA--KSGLENKTLAEYETKMAADLMAALQDMERLANSPSNHTEEVNIRK 162
 DB 1352 QEGKALQALSKANSEVAQMRKYETD-----ATQ-----RTELEAKK 1392
 OY 163 ALEAQDTDKLNTLVTONONKSLTEVLKTTDSADQIPAINSQ---LEINK---NSAD 215
 DB 1393 KI-AQR-----LQ-----DSEQVEAVNAKCAKSEKTKORLOGEYE 1427
 OY 216 OIKKLERONISYEAVLTNAGEVIRKASSEAGIKLGALQSIYDAGQSOA---AVLQAOQ 272
 DB 1428 DLKAVDERANSLAALADKKORNFQKLAEMKTKC-----BESQAELEALAKESR 1476
 OY 273 NNSPDNIA---ATKELIDAETKVN---LKOEHGTLDSPVKAKEQISOAKDIOEI 326
 DB 1477 SLSTELFKLNVEEALDQLETKRNKMLEQEIADLT-----EQIAENGKSIHEL 1527
 OY 327 KKS-----GSDIPVPGSSASAGSAGALKSSNSGRISLLIDVDNEN-AAIALQ 378
 DB 1528 EKSRKOMELEKADIDQWALEEALAEHEEAKIL-----RIQLELTQVKESEIDRIAKK 1580
 OY 379 GFRSMTEOPNVNPTAKILO--AMEAQLTAMSD-----QVYGADGELPAEIQAIIDALA 431
 DB 1581 D--EEIEQKRNKQRTVMOGLDEAVSRNEAIRLKKKMEGLNEIEIQSHANROA 1638
 OY 432 QALKOPSSADGLATAMGOVFAFAAKYGGSGAG-----TAGTOVMNVKOLYKTA 478
 DB 1639 EIKHK-----LRSVQGLKDTQHLDDALRGQEDLKEQLAIYERRANLLOAEVEEL----- 1689
 OY 479 FESSTSSSYAALSDGYSAYKTLNLSYSESRGVQSAISQIANPALSRSVSR----- 530
 DB 1690 -----RATLEQTERARKLAEOELDLSNERVQ--LHQTQNTSLIHTKKLETDLQOL 1738
 OY 531 -SGIESQGSASASORAAETIYRDSOTLDVYSRLQVLDLSMTSYVSNQANOEBIMOKL 589
 DB 1739 QSEVEDASDARNNAEKAKKAITDAMMAEELKEQ--DT--SAHLERKKMLEQIVKDL 1794
 OY 590 TASISAPQPGYAVONASDSLOKFAAOL--EREF--VDGERSLAESQENAFRKOPIAQOY 647
 DB 1795 QHRDEAEOL---ALKGKKQIQKLETRIRELEFELEGEQKNTESVCKLRYERRRVEL 1851
 RESULT 12
 TALA_DICDI STANDARD: PRT: 2492 AA.
 AC P54633:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Filopodin (Talin homolog).
 GN TALA.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxId=44689;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=AX2;
 RX MEDLINE=95213284; PubMed=7698984;
 RA Kretzmeier M., Gerisch G., Heizer C., Mueller-Taubenberger A.;
 RT "A talin homolog of Dictyostelium rapidly assembles at the leading
 edge of cells in response to chemottractant.";
 RL J. Cell Biol. 129:179-188(1995).
 RN [2]
 RP REVISIONS TO 2076; 2080; 2186 AND 2230.
 RA Mueller-Taubenberger A.;
 RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: ACTIN-BINDING PROTEIN THAT MAY BE INVOLVED IN THE
 CC CONTROL OF CELL MOTILITY AND CHEMOTAXIS.
 CC -I- SUBCELLULAR LOCATION: CYTOPLASMIC; RAPIDLY ASSEMBLES AT THE
 CC LEADING EDGE OF CELLS IN RESPONSE TO CHEMOTTRACTANT.
 CC -I- SIMILARITY: CONTRAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -I- SIMILARITY: IN THE C-TERMINUS TO YEAST SLA2 AND C.ELEGANS ZK370.3.
 CC PARTIAL, TO YEAST ROD1.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC
 CC EMBL: U14576; AAC46586.2; -
 DR DictyD; DD01121; TALA.
 DR InterPro; IPR000259; Band.4.1.
 DR InterPro; IPR002558; ILMEQ.
 DR Pfam; PF00373; Band.41; 1.
 DR Pfam; PF01608; ILMEQ.1.
 DR ProDom; PD011820; ILMEQ.1.
 DR SMART; SM00295; B41; 1.
 DR SMART; SM00307; ILMEQ.1.
 DR PROSITE; PS00660; BAND.41.1; 1.
 DR PROSITE; PS00661; BAND.41.2; 1.
 DR PROSITE; PS00507; BAND.41.3; 1.
 KW Structural protein; Cytoskeleton; Actin-binding.
 FT DOMAIN 134 294 BAND 4.1-LIKE.
 FT SEQUENCE 2492 AA; 268850 MW; 4B82F6BE6E7E752 CRC64;
 SQ
 Query Match 6.2%; Score 199; DB 1; Length 2492;
 Best Local Similarity 20.4%; Pred. No. 0.62; Indels 180; Gaps 28;
 Matches 147; Conservative 113; Mismatches 279; Indels 180; Gaps 28;
 OY 43 SEASPSSSVSSWFLSSAKNALIS---LRDAIINKS---SPTDLSOLEASTSTST 95
 DB 1831 STISMDIANYSIDIAHVSSAIIATTSQKIRSIITSSNOVIVSGDVIHKKV-STDK 1889
 OY 96 VTRVAKDYDEAKSNFTRAKSGLENK--TLAEY--ETKMADLMAALQDMF----- 142
 DB 1890 ANSSQAKVNDYRAVNDNTIRFLQSLKQALIGEILSDAIDQIRKVISDLQYSLFAAG 1949
 OY 143 RLANDSPNNHFEVNNIKKALEAKDPTDKINKLVTLQNO----- 183
 DB 1950 QLENDQSSOSTMNEVYTKQHLKNLQKDKTTLIVSSQLVGSSRCTQEHLSGATTKV 2009
 OY 184 -----NKSLEVLKTTDSADQIPAINSOLEINKNSADQIKDLERONISYEAVLT 233
 DB 2010 ANTVSSIVKTKANDIASVLADTTSSQDDILSASKALSISSQAWLAKKDAQRFKKDTA-FR 2068
 OY 234 NAGEVIKASSEAGIKLGALOS---IYDAG---DOSQAVLQAOQNNSPDNIAATREL 285
 DB 2069 SLGKSAEVAEAA---VGFELTSVYTAISDAGKIGIELKESIVQVANYHEKPTVLSNRDA 2125
 OY 286 I-----DAEFKVNLEQEHGTLNDSPLVYKKAKEQISQAOQDIEIRPSGSDPIV 336
 DB 2126 TAEITRQASRDIAKSSI-EIYTSYSSQDS-LVKSQSAVVSIVNVSFISSNK----- 2174
 OY 337 GPSGSAASAGSAAALKS-----SNNSGRISLLIDDV-----DNE 371
 DB 2175 ---GVIALLGNGNDLKKSVLENVKTQGDMLALLQCVKDDKNGSTSIADATRSISR 2231

(4)
RN X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 777-836.
RP MEDLINE-96419133; PubMed-8805510;
RA Houdusse A., Cohen C.;
RT "Structure of the regulatory domain of scallop myosin at 2-A
RL resolution: implications for regulation.";
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO F-ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY F-ACTIN.
CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X55714; CAA39247.1; -
DR PIR: S13557; S13557.
DR PIR: A40997; A40997.
DR PDB: 1SCM; 30-APR-94.
DR PDB: 1WDC; 11-JUL-96.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR000409; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; MYSC; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR MYOSIN: Muscle protein: Coiled coil; Thick filament: Actin-binding;
KW ATP-binding; Alkylation; Calmodulin-binding; 3D-structure.
FT DOMAIN 1 777
FT DOMAIN 1 778 805
FT DOMAIN 1 806 1338
FT DOMAIN 1 836 1938
FT NP_BIND 176 183
FT MOD_RES 693 693
FT MOD_RES 703 703
FT HELIX 778 821
FT TURN 822 823
FT HELIX 825 833
FT TURN 834 835
SQ SEQUENCE 1938 AA; 222821 MW; ASCE4127D1A4896 CRC64;
Query Match 6.1%; Score 197; DB 1; Length 1938;
Best Local Similarity 20.0%; Pred. No. 0.55;
Matches 148; Conservative 113; Mismatches 252; Indels 226; Gaps 29;
OY 5 HHMHSGPSSVSSNOSNMPT-----INGQT-----ASNSFKEST-----K 41
DB 1334 HADMDAIRQLDEEQESDVOROLSKANNEIQOWMSKPEEGSANTLEPELOOKRLLGK 1393
OY 42 ASEASPSSASSVSSWSSFLSKAKNALISLTDALINKNSPTDISLSQLEASTSTVTRVA 101
DB 1394 LSEADDTTAANAAKCSALKAKSRL-----QOELEDMSS-IEVRRANASVQOMK 1441
OY 102 KDYDEAKSNFTAKSGLENKTLAEVETKMDLMAALQDMERLANSPPNNHTEEVNNTK 161

DB 1442 K-----QRAFD-----KTTAEQAKVNSIQSELENSOK-----ESRGYSAELYRIK 1482
OY 162 KALEAQKPTIDKLNKLVTLQONOKSLTEVLKTTDSADQIPAINSOLEIKNSADQIKDL 221
DB 1483 ASIEEYQDSIG-----ALRRKKNL-----ADELHDLTDQLSEGRSTHELDKAR 1527
OY 222 ERONISYEAVLTVNAGEVINKASSEAGIKGLQALQISYVDADQSOAAVLQAO-----QNN 275
DB 1528 RLBEKE-----ELQALAEABGALQEGAEKAVMRQLEIATVRRNI 1569
OY 276 PNIATKELID-----AAETVNEIKQETHGLTDSPLVKKAEQ----- 315
DB 1570 DKRIEKEEDEFNTRRNHRALESQASLEAEAKGKADMRKIKKLEODINLEVALDAS 1629
OY 316 -----ISQAKDQIEIKPS-----GSDIPIVGGSSA-----ASA 345
DB 1630 NRGKAEKMTYKRYQOQIREMOTSIIEEORODEARESYNMARCTLMGSGVEELRAL 1689
OY 346 GSAGALKSSNNSGRISLLDDVD--NEMA--IALQFRSMIEQFNVPNPTAKELQAM 401
DB 1690 EQAERARKASDN-----LADANDVNEILTQSVSSVQGGKRL-----GDINAM 1734
OY 402 EAQLTAMSDQVADGELPAEIQAIKDALQALQKPSADGLATAMGQVFAAKVGGSSA 461
DB 1735 QTDLEMHGELKGAD-----ERCKKAMADAAR--LADELRAEDDH-----S 1773
OY 462 GTAGYQVMVNDLYTAPSTSSSYAALSDGYSATKLNLSYRSRGVQSATSQTN 521
DB 1774 NOYERVKRNLSEQVK-EFGIRLDEAEASSLKGKKMKIQLESVHELEELNE----- 1826
OY 522 PALSRVSRSGIESQGRSADASQRAAEFIVRDSQTLGVYSRLQVLSLMTSTIVSNQAN 581
DB 1827 -----QRRAEFTQKNMKADRLKE-----LAQADEDRKN 1857
OY 582 QE---EIMQKLTASISKAPQFGYPVAVNSDSLQNF-AAQLEREYVDSGLAESQENAF 637
DB 1858 QERLQELIDKLNKAIKTFKQYEEAEIYAINILAKYRKAKQHELEAEERADTADSTIQKF 1917
OY 638 R---KQAPFIQOVLVNIAS 653
DB 1918 RAKSSSVSVQSRSSVSVA 1936
RESULT 15
TFM1_HUMAN
ID TFM1_HUMAN STANDARD; PRT; 1093 AA.
AC P82094;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TATA element modulatory factor (TMF).
GN TMF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Cervical carcinoma;
RX MEDLINE-93028466; PubMed-1409643;
RA Garcia J.A., Ou S.-H.I., Wu F., Lusis A.J., Sparkes R.S., Gaynor R.B.;
RT "Cloning and chromosomal mapping of a human immunodeficiency virus 1
RT "TATA" element modulatory factor";
RL Proc. Natl. Acad. Sci. U.S.A. 89:9372-9376(1992).
CC -1- FUNCTION: THIS PROTEIN BINDS THE HIV-1 TATA ELEMENT AND INHIBITS
CC TRANSCRIPTIONAL ACTIVATION BY THE TATA-BINDING PROTEIN (TBP).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L01042; AAD54608.1; -
 DR TRANSFAC; T00835; -
 DR GeneW; HGNC:11870; TMF1.
 DR MIM; 601126; -
 KM Transcription regulation: DNA-binding; Repressor; Coiled coil.
 FT DOMAIN 439 922 COILED COIL (POTENTIAL).
 FT DOMAIN 984 1092 COILED COIL (POTENTIAL).
 SQ SEQUENCE 1093 AA; 123170 MW; 26133E8E5F4677BE CRC64; .

Query Match 6.1%; Score 194; DB 1; Length 1093;
 Best Local Similarity 19.2%; Pred. No. 0.36;
 Matches 145; Conservative 141; Mismatches 279; Indels 190; Gaps 28;

QY 10 SPSEYSSNOSNMPIINGQIASNS-ETKESTKSEASPSSASSSVSSFLSSAKNALIS 68
 DB 169 SSPKTEGKHHEVYNKESDKMVKPTVSLAKSESVIDVKTMTESISNTTQSTATKRIALE 228
 QY 69 LNDALINKSSPPLDLSQLEASTSTSTYVRAAKDYDE--ANSNPTAKSGLENAKTAE 126
 DB 229 PREQKHEDRQSMTPSPVPSTFSSGTSTTDIEYLDHESYSESSASSRQETDTSKSLHL 288
 QY 127 YEKKADLMAA-----LQDMERLANSPPNNHTEEVNNIK-KALEAQKDTIDKINKLV 178
 DB 289 MDTSFOLLSASACPEYNRLDDFOKLTESCSCSDAFERIDSFVSQSIDSR--SVSEINSD 346
 QY 179 TLONKSKLTVL-----KTTDSAD-----QIPAINSOLEINKNSADQIIRK 220
 DB 347 ELISGRKIALVPIIVNSSTPKTVESEBEGSEVENETLVIPTPEALMESGRATV--N 404
 QY 221 LERONISYPAVLTNAGEVIKASEAGIKLGQALQISTVDAQDSQA----- 265
 DB 405 CQPDILVSTPINEQTV-----LDKVAEQCEPAESQEPALSEKEDVCKTEFLIN 455
 QY 266 -----AVLQAQNNSPDNIAATKELIDAEKTKVELQEHGLDPL 308
 DB 456 EKLEKREAOQLLSKEKALEEFNLDKDEMRVKE-----ESSISLSLDEFT-----OR 506
 QY 309 VKKAEQI-----SQAKDIOEIKPSGSDPIVPGSGSASAGSAAGALKSNNGRT 361
 DB 507 IAEAEKVOACKERDAKKEIKNEE-----LATRLNSET 544
 QY 362 SLLDVDNEMAAIALQGRSMIEQFNVNPAATKELQAMEAQMTAMSDQLVGADGELPA 421
 DB 545 ADLKREKQIQRIGMEGKELSKQ--LHNSNITIKTRAKDKENEMVANKINKVELEE 602
 QY 422 ELQATDALAQAOKPSADGLATAMQVAFPAAKVGG--GSAGTGTQVMNKQLYKTA 478
 DB 603 ELQHLKQVL-----DGKEEVERKHRENIKKLNSWVEROEKDLGRLOVDDEL---- 649
 QY 479 FSTSSSYAALLDGYSAVKTLSLY-----SESRGVSQSAISO--TAMPALSRVSRSQ 532
 DB 650 --EKKRSTIOALD---SAVKELTLDHKANAAMKADSEQAERALSREKAKAEELSAALEKAO 704
 QY 533 ISSQ-----GRSADASQRAAEETVIRDSQTL---GDVYSRLQ-----VLDSI 570
 DB 705 EEARQOQETLAIQVGLRLALQRTQEAARKEDYLRHEIGELQQRLOQLEAMRNQDELISQV 764
 QY 571 MSIT-----VSNPQA-----NQEITMOKLTASISKAPQFGYPAVQNSADSLOKRAQDLERE 621
 DB 765 SSTTRPLRLQRIENTLQATLQSTSSWEKLEKNLS-----DRLGESOTLLAAVAVERE 814
 QY 622 FVDGERSLA-----ESQENAFRKQPAFIOQVL 648
 DB 815 RAATEELLANKTIOMSSMESONSLLROENSRFOQL 849

RESULT 16
 MYH3_HUMAN
 ID MYH3_HUMAN STANDARD; PRT; 1940 AA.
 AC P11055; Q15492;
 DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, fast skeletal muscle, embryonic (Muscle embryonic
 DE myosin heavy chain) (SMHC).
 GN MYH3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-89263803; PubMed-2726495;
 RA Eiler M.S., Steedman H.H., Sylvester J.E., Fertels S.H.,
 RA Rubinstein N.A., Kelly A.M., Sarkar S.;
 RT "Nucleotide sequence of full length human embryonic myosin heavy
 RT chain cDNA.";
 RT Nucleic Acids Res. 17:3591-3592(1989).
 RN [2]
 RP SEQUENCE OF 774-1940 FROM N.A.
 RX MEDLINE-90033298; PubMed-2806546;
 RA Eiler M.S., Steedman H.H., Sylvester J.E., Fertels S.H., Wu Q.-L.,
 RA Raychowdhury M.K., Rubinstein N.A., Kelly A.M., Sarkar S.;
 RT "Human embryonic myosin heavy chain cDNA. Interspecies sequence
 RT conservation of the myosin rod, chromosomal locus and isoform
 RT specific transcription of the gene.";
 RT FEBS Lett. 256:21-28(1989).
 RN [3]
 RP SEQUENCE OF 856-1940 FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE-90235862; PubMed-1691980;
 RA Bober E., Buchberger-Seidl A., Braun T., Singh S., Goede H.W.,
 RA Arnold H.H.;
 RT "Identification of three developmentally controlled isoforms of human
 RT myosin heavy chains.";
 RT Eur. J. Biochem. 189:55-65(1990).
 RN [4]
 RP SEQUENCE OF 856-1940 FROM N.A.
 RX MEDLINE-8936648; PubMed-2771643;
 RA Karach-Mizrahi I., Travis M., Blau H., Levinand L.A.;
 RT "Expression and DNA sequence analysis of a human embryonic skeletal
 RT muscle myosin heavy chain gene.";
 RT Nucleic Acids Res. 17:6167-6179(1989).
 RL
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- DEVELOPMENTAL STAGE: ABUNDANTLY PRESENT IN FETAL SKELETAL MUSCLE
 CC AND NOT PRESENT OR BARELY DETECTABLE IN HEART AND ADULT SKELETAL
 CC MUSCLE.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: X13988; CAA32167.1; -
 DR EMBL: X13100; CAA31492.1; -
 DR EMBL: X51593; CAA35942.1; -

```

DR EMBL: X15696; CAA3731.1; -.
DR PIR: S04090; S04090.
DR HSSP: P13538; 2MYS.
DR Genew: HGNC:7573; MYH3.
DR MIM: 160720; -.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811 IQ.
FT NP_BIND 840 1933 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 656 678 ACTIN-BINDING.
FT DOMAIN 758 772 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 696 696 ALKYLATION (SH-1).
FT MOD_RES 706 706 ALKYLATION (SH-2).
FT CONFLICT 1331 1331 A -> G (IN REF. 3).
FT CONFLICT 1391 1392 KK -> OE (IN REF. 1 AND 2).
FT CONFLICT 1608 1609 SR -> RA (IN REF. 3).
FT CONFLICT 1663 1664 RG -> QT (IN REF. 2).
SO SEQUENCE 1940 AA; 224035 MW; 43CA58C6A4BA1253 CRC64;

Query Match 6.1%; Score 194; DB 1; Length 1940;
Best Local Similarity 21.3%; Pred. No. 0.73;
Matches 155; Conservative 113; Mismatches 265; Indels 194; Gaps 34;

OY 8 MESGEPSVSSNOSSMNP-----INGQIA-----SNSETKES-----TKASEASPS 48
D 1232 LSSSMESVSKSKANLEKICRTLEDOLSEARKNEEIOHSLSLTTQKSLQTGEAGELSQ 1291
OY 49 ASSSVSSMPLSSAKNALI-----SLRDAIINKNSPTDLSLEASTSTSTYTRVANDY 104
D 1292 LEEKESIYSQSRSKQAFTQOTEELKROLEEENKAKMALAHALQSSRHDCDLREQVEE 1351
OY 105 DEAKSNFDTA--KSGLENAKTLAEYETKMDLALQDMERLANSDPNHNTEEVNNIK 162
D 1352 QEGKAEIQRAISKANSEVAQWRTKETD-----AIQ-----RTEELEAKK 1392
OY 163 ALEAQKDTIDKLKLVTTQONKSLTEVLKTTDSADQIPAINSQ---LEINK-----NSAD 215
D 1393 KL-AQR-----LQ-----DSEEOVEAVNAKCASELEKTKQRLQGEYE 1427
OY 216 QIKKLERONT-----SYEAVLT---NAGEVITASSAGIKIGQALQSIYDADGQ 262
D 1428 DLMVDVEPANSIALAALDKKORNFEDVLAEMWKCEESQELDASIKESLSL----- 1478
OY 263 SOAAVLIAQOONNSPDNIATKELIDAEETKVE---LKOHTGLTDSPLVKAQEIOISA 319
D 1479 --STELFKLN-----AYEALDQLETVKREKNKLNLEGEIDL-----QIAGEN 1520
OY 320 QKDIOEIRPS-----GSDITVPPSGSASAGSAGALKSSNNGRISLLDDVDNEM 372
D 1521 GRTHELKSRKQIELEKADIQIALEEAALEHEEAKTL-----RIQLELTQVKSSEI 1573
OY 373 -AAILOGFRSMIEQFNVNPNPATKELO-AMEAQLTAMSD-----OLVAGDGLPAEIQ 424
D 1574 DRKIIEKD--EETQELKRNRYQTTVETMOSALDAEVRSRNEAIRLKKMKGGDINEIETQS 1631
OY 425 AIKDIALQALKOPSDGLATAMGOVAFAAKVGGGSAG-----TAGTVOMNV 471

```

```

DB 1632 HANROAEFLKH-----LRVGGOLKDPQLHLDALRCQEDLKEQLATVERRANLQAEV 1686
OY 472 KOLYTAFTSSSSSSYAAALSDGYSAYTTLNSLYSESSGVSQSAISQTAANPLASVSFR- 530
D 1687 EEL-----RATLEOTERARKLAEOELDSNERVO--LLHQNSTLIHTKKKL 1731
OY 531 -----SGIESQGRSADASORAEITVRDSQTLGDVYSRLQVLDLSMTSTVSPQANO 582
D 1732 ETDLMQLOSEVEDASRDARNNEEKAKKAITDAMMAEELKEQ--DT--SAHLEMKKNL 1787
OY 583 EIMQKLTASISKAPQFCYPAVQNSADSLQKFAQL-EREF-VQGESLAESQENAFKQ 640
D 1788 EGTVDLQHRDEAEQL--ALKGKKQIQKLETRIREFELFELEEQKKNESVAGLKRY 1844
OY 641 PAFIOOV 647
D 1845 ERRVKEL 1851

RESULT 17
MISA_DROME STANDARD; PRT; 1962 AA.
AC P05661;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin heavy chain, muscle.
GN MHC.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC STRAIN=Canton-S; TISSUE=Pupae;
RX MEDLINE=89384556; PubMed=2506434;
RA George E.L., Ober M.B., Emerson C.P., Jr.;
RT "Functional domains of the Drosophila melanogaster muscle myosin
RT heavy-chain gene are encoded by alternatively spliced exons.";
RL Mol. Cell. Biol. 9:2957-2974(1989).
RN [2]
RP SEQUENCE OF 1-264 FROM N.A.
RX MEDLINE=87280141; PubMed=3038896;
RA Wassenberg D.R., II, Kronert W.A., O'Donnell P.T., Bernstein S.I.;
RT "Analysis of the 5' end of the Drosophila muscle myosin heavy chain
RT gene. Alternatively spliced transcripts initiate at a single site and
RT intron locations are conserved compared to myosin genes of other
RT organisms.";
RL J Biol. Chem. 262:10741-10747(1987).
RN [3]
RP SEQUENCE OF 333-614 FROM N.A.
RC STRAIN=Canton-S; TISSUE=Embryonic muscle;
RX MEDLINE=91330870; PubMed=1907912;
RA Kronert W.A., Edwards K.A., Roche E.S., Wells L., Bernstein S.I.;
RT "Muscle-specific accumulation of Drosophila myosin heavy chains: a
RT splicing mutation in an alternative exon results in an isoform
RT substitution.";
RL EMBO J. 10:2479-2488(1991).
RN [4]
RP FUNCTION: MUSCLE CONTRACTION.
CC -1 SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1 SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1 ALTERNATIVE PRODUCTS: At least 15 isoforms: AAAA (shown here),
CC BBDBA, BABDB, 3b, 7b, 7c, 7d, 9c, 11b, 11c, 11d, 11e, 15b and
CC 18; may be produced by alternative splicing. Exons 3, 7, 9, 11 and
CC 15 are mutually exclusive splicing exons and exon 18 is included
CC or excluded
CC -1 TISSUE SPECIFICITY: Expressed in larval and adult muscles.
CC Isoforms containing exon 9a are expressed in indirect flight
CC muscles, exons 9a and 9b are expressed in jump muscles, exons 9b

```


OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RC MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Squires J., Peat N., Hayles R., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jaegels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Murphy L., Niblett D., Odell C.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynponez B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinhardt H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard J., Tallada V.A., Garçon A., Thode G.,
 RA Dague R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Shkurtel L., Lowe T., McCombie W.R., Paulsen I., Petashkin J.,
 RA Shkurtel L., Lowe T., McCombie W.R., Paulsen I., Petashkin J.,
 RA "The genome sequence of Schizosaccharomyces pombe".
 RA Nature 415:871-880(2002).
 RN [2]
 RP SEQUENCE OF 495-1727 FROM N.A. AND CHARACTERIZATION.
 RC STRAIN=972;
 RC MEDLINE=20123449; PubMed=10660053;
 RA Jimenez M., Petit T., Gancedo C., Goday C.;
 RT "The amlt gene from Schizosaccharomyces pombe encodes a coiled-coil
 RT protein that associates with the medial region during mitosis".
 RL Mol. Gen. Genet. 262:921-930(2000).
 RN [3]
 RP SEQUENCE OF 644-834 FROM N.A.
 RC STRAIN=968 h90;
 RC MEDLINE=20223868; PubMed=10759889;
 RA Ding D.O., Tomlita Y., Yamamoto A., Chikashige Y., Harguchi T.,
 RA Hiroka Y.;
 RT "Large-scale screening of intracellular protein localization in living
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library".
 RL Genes Cells 5:169-190(2000).
 CC -1- FUNCTION: AFTER THE ONSET OF MITOSIS, AT MID- TO LATE ANAPHASE,
 CC CO-LOCALIZES WITH THE MEDIAL ACTIN RING. MAY PLAY A ROLE IN
 CC CYTOKINESIS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AL133357; CAB62414.1; -;
 DR EMBL: AF010473; BAB65416.1; ALT_INT.
 DR EMBL: AB028012; BAA87316.1; -;
 KW Coiled coil.
 FT DOMAIN 57 361 COILED COIL (POTENTIAL).
 FT DOMAIN 443 463 COILED COIL (POTENTIAL).
 FT DOMAIN 542 740 COILED COIL (POTENTIAL).
 FT DOMAIN 804 1106 COILED COIL (POTENTIAL).
 FT DOMAIN 1223 1427 COILED COIL (POTENTIAL).
 FT DOMAIN

FT DOMAIN 1497 1555 COILED COIL (POTENTIAL).
 FT DOMAIN 1601 1664 COILED COIL (POTENTIAL).
 SO SEQUENCE 1727 AA; 197858 MW; F820BF8D9C132644 CRC64;
 Query Match 6.0%; Score 192.5; DB 1; Length 1727;
 Best Local Similarity 19.6%; Pred. No. 0.73;
 Matches 140; Conservative 131; Mismatches 286; Indels 157; Gaps 29;
 OY 13 EASVSSNOSMPLINQOINSM-SEFKEPSKASEASPSASSVSSWFLSAAKNALISLTD 71
 DB 101 ESKITLLENQNRFPDALTENOSLRANSELQDSIASDQ-----LSIADQI---E 150
 OY 72 AILNKS-----SPTDSLQLE-----ASTSTVTYVAAKDYDEAKSNPTAKS 116
 DB 151 ALONENSHLGEVQASHQALSDIEERKKQHPMASSSRVKEEILVQEKALVSALVSLAS 210
 OY 117 G-----LE-NAKTLAEETKMDALQDM-ELANSDPSNNHTEVNNIKKALEQ 167
 DB 211 DSKVCEKLEVSRRQVQDLEKLAGLAQONTLENERIQLEQKRNSVSDGINSKILETD 270
 OY 168 KPTIDKRLVTLQONKSLTEVLKTTDSADQIPAINSQL-----EINKNSAD 215
 DB 271 PSIKLEEEVEFQKRLTALME-SKSELSQSEVVALQEKLTQOQSILYNNVTELNKKQ 329
 OY 216 QIIKD--LERONISYEAVLTNAGEVIKASSEGIKGLQALQSLVDAGDSQAAYVLAQON 273
 DB 330 LIISENSELRELQEKYVSSEL-QVYKENKNTSVSAGVGLFSP-----AKLSAVON 381
 OY 274 NSPDNIATKELID--AAETYNELQKHTGLTD--SPVYKKAERQISOAQDIOIKRS 329
 DB 382 ---PESEFTRKVSVDNKKLQKVSLSKLQDRLTRKFSFCEQVKQIPVAKQDRSEIVN 438
 OY 330 GSDIPVPGSGSASAGSAAGLKSSNSNGRISLILDVDNEMAAITAGLFRSMIEQFNV 389
 DB 439 NIYMNFLSES-----LETSSNN-----LTQVQAEILSTKMN----- 469
 OY 390 NNPAKLELQMEAOITLAMSOLVGADGL-----PAETQAIKDALQALQKPS 438
 DB 470 -DEACYLQTLASTQCSDSLREVICLMAELDLNETKSHNVPAVVALDEVAQ--NPS 525
 OY 439 A-----DELATMGQVAFPAAKVGGSGAGT--AGYOVMMNKLYKTAFSST 482
 DB 526 TASETVLNKELANFSSIKKAVSKTLELRKVRALBECDEVLQOTVOYQISNAVKE--NSN 583
 OY 483 SSSVAAALSDGYSAYKNTLSYSESRSQVQSAISOTANPALRSVRSRGISQGRSADA 542
 DB 584 TLSEQIKNLESELNKSINKINESLNRNLKELANRSRSLISHNSAGNIDKMSIDE 643
 OY 543 SORPAETIVRDSOTLGDVTSRLQVLDLSMSTIVSNPQANQDEIMQKLTASISKAPOFGTP 602
 DB 644 STRELE-----KNEYEVNEMTALIOESLSKRNDLISEMA-IRK----- 682
 OY 603 AVONASDLSQKFAQLEREFEVGERSLASQENAFKQAPFQOVVNLASLFS 656
 DB 683 ELNLS-----KITQOQLSTRLTN---ANNVDYAEFKKEKELRSINQNLDTITS 727
 RESULT 20
 MYH4_RABIT
 ID MYH4_RABIT STANDARD; PRT; 1938 AA.
 AC Q28641;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE 15-YOUN-2002 (Rel. 41; Last annotation update)
 OS Myosin heavy chain, skeletal muscle, juvenile.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand white; TISSUE=Skeletal muscle;
 RA Maeda K., Hostinova E., Roesch-Kleinhauf A., Schuster H., Gasperik J.,


```

RA Wittinghofer A.;
RT "Isolation, sequencing of myosin heavy chain cDNA from rabbit
RT skeletal muscle and a novel cosynthesis of S-1 fragment with the
RT essential and regulatory light chains."
RL Submitted (Aug-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32574; AAA74199.1; -.
DR HSSP: P13538; 2MYS.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR002928; Myosin_tail.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF01576; Myosin_tail; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR MYOSIN: Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 783 MYOSIN HEAD-LIKE.
FT DOMAIN 1 784 813 IQ.
FT DOMAIN 1 842 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 658 680 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 760 774 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 35 35 METHYLATION (MONO-) (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 552 552 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 756 756 METHYLATION (TRI-) (BY SIMILARITY).
FT MOD_RES 698 698 ALKYLATION (SH-1) (BY SIMILARITY).
FT MOD_RES 708 708 ALKYLATION (SH-2) (BY SIMILARITY).
SQ SEQUENCE 1938 AA; 223064 MW; D8A8A2EC5B182626 CRC64;

Query Match 6.08; Score 192; DB 1; Length 1938;
Best Local Similarity 19.68; Pred. No. 0.88; Mismatches 227; Indels 122; Gaps 22;
Matches 111; Conservative 107;

OY 102 KDYEAKNPTFANSGLLENATLT-AEYETKMAADLMAALQDMERLANSPPSNHTEEVNNI 160
DB 850 KEMANMKKEEFKTESLAKAEKKELEKKNVALM-----QEKNDL 890
OY 161 KKALEAKODT-----DKMKLVTLQONKSKLSEVTKTDSADQIPAINSOLEINKNSADO 216
DB 891 QLOVOALDSLADAEERCDQLTKTKIQLEA-KIKREVERAEDEEIEINAEITAKKRKLEED 948
OY 217 IKDLERONISYEAVLTNAGEVIRKASSEAGIK-----LGOALQSLVDAGDSQAAVL 268

```

```

DB 949 ECSEELKIDDELLETLAKV-EKEKHATENKVNLTREMGADERTYAKLT-----KCKKALQ 1003
OY 269 QAOQNSNDPNIATATELLDAETKYNELKQEHTEGTLDSPLVYKAEQISOAKQDOELKP 328
DB 1004 EAHQOTLDD-----LQAEEDKYNTLTAKT-----KLQOQVDDLEGSLEQEK 1046
OY 329 SGSDPIVPGSGSASAGALKSSNNSGRISLLDDVDNEMAAIALOGFRSMIEDFN 388
DB 1047 IRMDL-----EAKKKLEBDLKLAEOSTY-----DIENDKQDLBKLRKKEFESN 1092
OY 389 VNNPATAKELAMEAQLTAMSDOLVAGDELPAEIOAIKDALQALOPASDGLATAMQO 448
DB 1093 LQ--SKIDEQALAMQLOKIKTEIQARIEELEEBEIEARASRAKAKORS--DISRELEE 1148
OY 449 VAFAAKVGSGSAGTAGYQNMVNVKQLYKTAFSTSSSSYAAALSIGYSAVYKTLNSLYSES 508
DB 1149 ISERLKEAGG--ATSAQIEENKRR-----EAEFQKM 1177
OY 509 RSGVQSA-ISOATNPALRSVSRSGIESQGRSADASORAETIYRDSOTLGQVYSRLQVL 567
DB 1178 RQDEEATLQHEATATATLRKHADSVALGEOIDMQLVKOKLEKESL-----KMEI- 1231
OY 568 DSLNSTI--VSNPQANOEIMOKLTPASIS--KAPQGYPAVQNSADSLOKFAQLEREFTV 623
DB 1232 DDLASNMETVSMAKGNLEKMCRTLEDQVSELKTRKEEH---QRLINDLSAQRARLQTESG 1288
OY 624 DGERSLAESQ---ENAFKRQPAFTQOV 647
DB 1289 EFSRQDEKDSLVSQLSRGAFTQOI 1315

RESULT 21
ANT1_ONCVO STANDARD; PRT; 2022 AA.
ID ANT1_ONCVO
AC P21249;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Major antigen.
GN OVT1.
OS Onchocerca volvulus.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Splirurida; Filarioidae;
OC Onchocercidae; Onchocerca.
OX NCBI_TaxID=6282;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95287898; Pubmed=7770081;
RA Tritreapapab S., Ritchie T.L., Tuan R.S., Shepley K.J., Dلمان J.D.,
RA Neubert T.A., Scott A.L.,
RT "Molecular cloning of a gene expressed during early embryonic
RT development in Onchocerca volvulus."
RL Mol. Biochem. Parasitol. 69:161-171(1995).
RN [2]
RP SEQUENCE OF 733-866 FROM N.A.
RX MEDLINE=89127417; Pubmed=2464764;
RA Donelson J.E., Duke B.O.L., Moser D., Zeng W., Etoudu N.E.,
RA Lucius R., Renz A., Karam M., Flores G.Z.;
RT "Construction of Onchocerca volvulus cDNA libraries and partial
RT characterization of the cDNA for a major antigen."
RL Mol. Biochem. Parasitol. 31:241-250(1988).
CC -1- FUNCTION: MAY BE A MYOFIBRILLAR PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U12681; AAA80009.1; -.
DR EMBL: J03995; AAA29412.1; -.

```


QY 159 NIKKALEAKDTI---DKLNKLVTLQONKSLTEVLTSTDSADQIPAINSOLEINKNSA 214
 Db 890 DILQLOQAEADSLADAEERDQILKTKIQLDA---KIKETERADEEINAEILTKARKRL 947
 QY 215 DQIKLEKRONISYEAVLTFNAGEVITKASSPAGIK-----LGOALQSTVDAGDOSQA 266
 Db 948 EDCSCSELKADIDLELTLAKV-EKEKHATENKKNITTEEMAGIDETIAKIT---KEKKA 1002
 QY 267 VLOAONNSPDNIATKRELIDAEKTVNELKQEHGTLTDSPLVKAAEQISQAOKDIOEI 326
 Db 1003 LQAHQOITLDD-----LQAEEDKVNLTTLAKI-----KLEQVVDLEGSLEGE 1045
 QY 327 KPGSGPIPIVPGSASAGSAGAAKSSNNSGRISILLDDVDNEMAAI-----AL 377
 Db 1046 KKTIRMDL-----BRARKKLEGGDKLQOESAM-----DIPNDQOOLDEKTKKEFEF 1091
 QY 378 QGFRSMIEQFNVNPPATKELQAMEAQLTAMSDQVAGDELPAEIOATKDALQALOKP 437
 Db 1092 SGIQSKIE-----DEQALGMLOLQKIKELQARITELQAELEEEIEAERASAKAEKOR 1140
 QY 438 SADGLATMAGQVFAAAKVGSGSAGTAGTVQNNVQK-----LYKTAFFSST 482
 Db 1141 S-DLSRELEIEISERLEEAGG---ATSAQIEMNKKREAEFQKMRDLQEAATLQHEATPAAT 1195
 QY 483 SSSSYA---AALSDGYSAKTYLNLSEKSSGVQSAISOTANALSRYSRSIEGSGRS 539
 Db 1196 LRRKHADVAELGEQIDNLRQVKQLEKESEKMEIDDLASMETVSKAKGMLEKMCRA 1255
 QY 540 AD-----ASQRAETIVD-----SOTLDGVSR-LQVLDSTMTSVSNPOA---N 581
 Db 1256 LEQQLSEIKTEKEEQRLINDLTAQARARLOTESGETSRQDEDEDTLSQLSRCKQAFQO 1315
 QY 582 QEEIMOKLTASISKAPQFQYPAVONSADSLQKFAQLEREFEVDGERSLAESQ 633
 Db 1316 IEELKROLEEBE-KAKSALAHALQSSRHCDLRLQEQYEE---QEKAKELQ 1362

RESULT 24
 MYSD_CAEEL
 ID MYSD_CAEEL STANDARD; PRT; 1938 AA.
 AC P02567; Q19674;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain D (MHC D).
 GN MYO-1 OR R06C7.10.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RX NCBI_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA MEDLINE=89178677; PubMed=2926820;
 RA Dlib N.J., Maruyama I.N., Krause M., Karn J.;
 RT "Sequence analysis of the complete Caenorhabditis elegans myosin
 heavy chain gene family.";
 RT J. Mol. Biol. 205:603-613(1989).
 RN [12]
 RP SEQUENCE OF 34-1795 FROM N.A.
 RX MEDLINE=83273600; PubMed=6576334;
 RA Karn J., Brenner S., Barnett L.;
 RT "Protein structural domains in the Caenorhabditis elegans unc-54
 myosin heavy chain gene are not separated by introns.";
 RT Proc. Natl. Acad. Sci. U.S.A. 80:4253-4257(1983).
 RN [3]
 RP SEQUENCE OF 115-365 AND 1492-1763 FROM N.A.
 RX MEDLINE=85201409; PubMed=3888374;
 RA Karn J., Dlib N.J., Miller D.M.;
 RT "Cloning nematode myosin genes.";
 RT Cell Muscle Motil. 6:185-237(1985).
 RN [4]
 RP SEQUENCE FROM N.A.

RC STRAIN-Bristol N2.
 RA Gardner A., McMurtry A.;
 RL Submitted (Apr-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
 C.ELEGANS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 DR EMBL: X08065; CAA30854.1; -;
 DR EMBL: M37232; AAA28119.1; -;
 DR EMBL: M37234; AAA28120.1; -;
 DR EMBL: Z71266; CAA95848.1; -;
 DR EMBL: Z71261; CAA95848.1; JOINED.
 DR EMBL: Z71261; CAA95806.1; -;
 DR EMBL: Z71266; CAA95806.1; JOINED.
 DR PIR: S02772; MKWL.
 DR HSSP: P08799; LMND.
 DR WormBep: R06C7.10; CE06253.
 DR InterPro: IPR004009; Myosin.N.
 DR InterPro: IPR002928; Myosin.tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF01576; myosin_tail; 1.
 DR Pfam: PF02736; Myosin N; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head; 1.
 DR SMART: SM00242; MSC; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 FT DOMAIN 1 845
 FT DOMAIN 846 1938
 FT DOMAIN 846 1170
 FT DOMAIN 1171 1938
 FT NP_BIND 177 184
 FT DOMAIN 660 682
 FT DOMAIN 764 778
 FT MOD_RES 128 128
 FT MOD_RES 700 700
 FT MOD_RES 710 710
 FT MOD_RES 94 94
 FT CONFLICT 98 98
 FT CONFLICT 377 377
 FT CONFLICT 377 377
 FT CONFLICT 389 380
 FT CONFLICT 391 391
 FT CONFLICT 408 408
 FT CONFLICT 474 474
 FT CONFLICT 577 577
 FT CONFLICT 681 681
 FT CONFLICT 1373 1373
 FT CONFLICT 1659 1659
 DV -> D (IN REF. 2).
 W -> N (IN REF. 4).
 W -> N (IN REF. 2).
 Q -> G (IN REF. 2).
 L -> F (IN REF. 4).
 I -> N (IN REF. 4).
 S -> D (IN REF. 2).
 E -> Q (IN REF. 3).

D b	307	NTSYDKATAAEVMSKNNVNDCT--SSEHSHAG-SNMSKINASQ-----	350
Q y	329	SGSDIPYVPGSGAASAGSAGALKSSNNSGRT-SLLDDVDNEMAIALO---GFRSM	383
D b	351	-SSDLMSGPGN-----IKSHSTSKNTDYNVALDENAGSAGSISBQIGKNGRSL	398
Q y	384	IE-----QNVNVPATAKELOMEAOLTMASQLYGAGELPAE-----IQAIKDA	429
D b	399	NESSIESGRKAESRNNTAADLTDSDVANGTVSSSHSKSASGCTSLDENHNKTHALQASVDE	458
Q y	430	LAQALKOPASAGLATPAMQOVFAAKKGGGSGAGTGV-OMNYYKQLYKTAFSSSTSSSYA	488
D b	459	HGN-WKHNISIDGSTRNKTGTFEGFSGSEBASIKKADGTIMSOVYNK-----NOTNRTYE	510
Q y	489	A---ALSDGYAYVTLNSLYSESRGVOASISQIOT--ANPAL-SRYSRSGIESOGRSAD	541
D b	511	AEKSALEKHNHE--KNSDGTFEDESKSGNSRVNRTDGGSNLNAVGSYGVKRGVSSNETIAS	568
Q y	542	AS-----QRAAE-----TIYRDSQTLDDVYSRLQVLDLSLSTIYSPQAN	581
D b	569	SNAPRTSDAESNOTDHLHQKTANGTELTTHAKDSQOVAASANANSSLDTSMSAA--DAKGN	626
Q y	582	QEEIMOKLTASISKAPQFGYPAVONASDASLOKFAAQLER-EFYDGEKS-LAESQENAFRK	639
D b	627	-----KYDKTSQQAAD-SHDAISASSDDYAKIYKHADRBSESISNDSNQTASAEHNDSSK	679
Q y	640	Q 640	
D b	680	Q 680	

RESULT 27	
YIQ9_YEAST	
ID YIQ9_YEAST	STANDARD;
	PRT; 995 AA.

DT		01-FEB-1995 (Rel. 31, Created)
DT		01-FEB-1995 (Rel. 31, Last sequence update)
DT		15-JUN-2002 (Rel. 41, Last annotation update)
DE		Hypothetical 99.7 kDa protein in SDL1 5'region precursor.
GN		YIL169C OR Y19402.07C.
OS		Saccharomyces cerevisiae (Baker's yeast).
OC		Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC		Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX		NCBI_TaxID=4932;
RN	[1]	
RP		SEQUENCE FROM N.A.
RC		SPRAIN-S288C / AB972;
RA		Barrall B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA		Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,
RA		Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagsels K., Jones M.,
RA		Louis E., Lyne G., Moule S., Moule T., Odell C., Pearson V.,
RA		Rajandream M.A., Riles L., Rowley N., Skelton J., Smith D.,
RL		Walsh S.V., Whitehead S.;
RT		Submitted (DEC1994) to the EMBL/Genbank/DDBJ databases.
CC		
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (See http://www.isb-sib.ch/announce/ ,
CC		or send an email to license@isb-sib.ch)
CC		
EMBL:	Z46921	CAB87023.1; .-
DR	SGD:	S0001431; YIL169C.
DR	InterPro:	IPIR004089; Chtaxis_t1rnsnd.
DR	InterPro:	IPIR000727; T_-SNARE.
DR	PROSITE:	PS50192; T_SNAKE: UNKNOWN_1.
KW		Hypothetical protein: Signal.
FT	SIGNAL	1 23 POTENTIAL.
FT	CHAIN	24 995 HYPOTHETICAL PROTEIN YIL169C.
FT	DOMAIN	92 154 T-SNARE COILED-COIL HOMOLOG (POTENTIAL).
FT	CARBOHYD	28 28 N-LINKED (GLCNAC...) (POTENTIAL).

FT	CARBOHD	35	35	N-LINKED (GLCNAC. . .)	(POTENTIAL.)
FT	CARBOHD	468	468	N-LINKED (GLCNAC. . .)	(POTENTIAL.)
FT	CARBOHD	664	664	N-LINKED (GLCNAC. . .)	(POTENTIAL.)
SQ	SEQUENCE	995 AA;	99735 MW;	F63E287A03F137EC	CR64;

Query Match	5.98;	Score 188.5;	DB 1,	Length 995;
Best Local Similarity	20.98;	Pred. NO. 0.54;		
Matches 142;	Conservative 119;	Mismatches 274;	Indels 143;	Gaps 27;

Oy	10	SGPEVSSVSSNOSMNNI-----INCQIANS-----ETMESTR--ASEMSPS	48
Db	27	SNSTSISSSSSSTSVSSSSSGSVSISISSIATSSSADTILSSITQSAASSTGVSSSVGPS	86
Oy	49	ASSSVSSAFESASAKNALISLRDAIILKNNSEPTD--SLSOLEASTS--STSTVRVAKD	103
Db	87	SSSVYS--SSVQSSSSSVSDVSSSVSSQSSSSASADVSSSSQASSTIDVSSSVQSSSSA	144
Oy	104	YDEAKSNFDJARSGLENAKTLAEYETKMAEDLMAALQDMERLANSDDPSNNHTEEVNNIKKA	163
Db	145	SDVSSSVQSSSSASADVSSSVQSSASASADVSSSV--SQSASSTSDVSSSVQSSSSASDV	203
Oy	164	LEAQKOTIDKLKLVLTQONQKSLTEVLEKTTDSADQIPLAINSOLEIKNNADQIIDLER	223
Db	204	SSSVSSQSSSSASADVSSSVQSSAS--STSDVSSSVQSSASSTSGVSSSGSOVSASAG	258
Oy	224	QNISVEAVLTNAGEVTRKASSSAGIKIKLOALOSIYDADQOQAIVLQAOQNNSPD-----	277
Db	239	SSSSSEPOSTSS-----ASTASGASTNSISLSTSSA--SSASATVANSLSSSDGTIYLP	310
Oy	278	-----NIAATKELIDAEFTKYNELKOEHTGLT-----DSPLKKAKEQISQA	319
Db	311	TTTTISGDLTLTGKVIATGTGGVVAAGAKITLLDGKISFSADLKYGGDLLYKKSRETYPGT	370
Oy	320	QKDIQETKPGSSDPIPIVPGSSASASAGSAAACAL--KSSNNSGRISILL-----	365
Db	371	EFDI-----SGENPDYTCGNFAEESAATSAISYTFPTPSFPDNGSDISSLSSKSKKEVTF	425
Oy	366	DDVDNEMA-----AIALOGFISMIE-----QFVNNPATKELQAMEAQLT--	406
Db	426	SPYSNSGAFSPFSNAILNGSGVSGIQRDDTEGSVNNGEIMLDNGSTVIYIEPVGKGTVN	485
Oy	407	-----ANSDQVGADEGLPEAICQAIKALQALQKPSADSLATAMQVAFMAA	434
Db	486	IISGNLYAHYDPTTGGTGVFVFKGE--GVLAVDPTETNAPPIPVG--YTGKNQAIATTAO	540
Oy	455	KVGGGSACTAG--TYQMMVKOL--YKTFASSTSSSYAALSDGYSAYKTLNLSYSESR	509
Db	541	ITALSYDGTGLVLTATGQNRQPSAIGTGSSSPFSVSEGIKACATAYIYLYNKGAVATS-	599
Oy	510	SGVQSAISQOTANPALSRVSRS--GIESQGRSADASQARAETTVRDSQTLG-----D	559
Db	600	-----AASSSTASGA-SASVYTGSTSFGASVYGTASTSFGASVYGTASTSFGASVYGTST	654
Oy	560	VY-SRLQVLDLSLMTIYS	576
Db	655	VYTTTLDYVNAFSTVYVVS	672
RESULT 28			
MYSN_DROME STANDARD; PRT; 2017 AA.			
AC	099323;		
DT	01-JUN-1994 (Rel. 29, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DE	Myosin heavy chain, non-muscle (Zipper protein) (Myosin II).		
GN	ZIP.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;		
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
NCBI	taxid=7227;		
LN	[1]		

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X57841; CAA40973.1; -
DR EMBL: M38210; AAA26977.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR ProDom: PD153432; Csurface_antigen.1.
KW PROSITE: PS50847; GRAM_POS_ANCHORING.1.
KW Antigen; Cell wall; Peptidoglycan_anchor; Repeat; signal.
FT SIGNAL 1 50 POTENTIAL.
FT CHAIN 1 51 CELL SURFACE ANTIGEN 1.
FT PROPEP 1507 1506 REMOVED BY SORTASE (POTENTIAL).
FT SITE 1503 1507 LEPTG SORTING SIGNAL (POTENTIAL).
FT MOD_RES 1506 1506 AMIDE-LINKED TO CELL WALL (POTENTIAL).
FT CONFLICT 427 427 A -> E (IN REF. 2).
FT CONFLICT 431 431 Q -> K (IN REF. 2).
FT CONFLICT 434 434 Q -> S (IN REF. 2).
FT CONFLICT 531 531 A -> S (IN REF. 2).
FT CONFLICT 600 600 A -> S (IN REF. 2).
SQ SEQUENCE 1528 AA; 165572 MW; B235F9CCD92838E0 CRC64;

Query Match 5.9%; Score 188; DB 1; Length 1528;
Best Local Similarity 21.3%; Pred. No. 0.96;
Matches 159; Conservative 107; Mismatches 291; Indels 190; Gaps 32;

9 ECGPESVSSNOSSMNPITNGQIASNETSEKSTKASAPS-----ASSSSWSGF 58
DB 52 EISTSTSGDTRAVVGTENGPNATNLPDKNDPSSQAEITSOAQARKTAMSDVSTSEL 111
QY 59 ISSARN---ALISLDAIANKNS-SPTDSLQLEAS---TSTSTVTRVAKDY--DE 106
DB 112 DEAAKSPQAGTVSODAVNKGTVPESDANQKEPEIKDYSKQAADIQKATEDYKASV 171
QY 107 AKSNEDTAKSGLENKATLAEYETKKAADLMAALQDMR-----LANSDPEN 151
DB 172 AANQGETDINOEIAKKKQYE---ODLAANKAEVERSLMRKRPPIYEAKLAQNKQDL 228
QY 152 NITEEVN-NIKKALEOKPTIDKLKLVTLQN--QNKSLTEVL-KTSDSADQIPAINSQL 207
DB 229 AAIQOQNSQAPAAKAYDEKARVQAAANAAKAEALAAANTAKNDQIK--EI 285
QY 208 E-INKNASADIIKLERONISYEAVLTNGEVIKASSEAGIKLQA-----LQSIVDAGD 261
DB 286 EAIQGRSAK---ADYEAKLAQYEKDLA-AAQAGMANANEADYQAKKAAVEQELARVQANA 341
QY 262 GSOAAVLAQOQNNSPDNIAATKELIDAEFTKYNELKOEHTGLTDFLVKKA--EQDISQA 319
DB 342 AKKOAYEQALANSAKNAQITAE-----NEAIQOQNA-----QAKAYEAKLAQY 385
QY 320 OKDIOEIKRSGSDIPVSGSAGSAAGALKSSNNSGRISLLIDVDNEMA----- 373
DB 386 QKDL-----AAQSGNNA-----NEADYQEKLAAYEKELEARVQOAN 421
QY 374 AIALOGFRSMIOFNVN-----PATAKELQAMEAQLTANSDQLVGAD--GEL 419
DB 422 AAKQAQVEQVOQANAKNAKTEANRAIRERNAKAKTDELKSKYQOEELAQKKDLAY 481
QY 420 PAEIQ-----AIDALAQALKQPSADGLATAMGOVAAAKAVGGSGAGTACTVOMNK 472
DB 482 PAKLAYODEQAIRKAALELEKHKMED-----GNLSPEPSASQSLVYDLE 525
QY 473 QLYKTAFTSSSSYAALSDGSAYK-----TINSLSYSSRSGVQSAISQTN 521
DB 526 PAAQVALVYDGLKIRKASALDEAFSHDEKYNHNLLOPNLNTYILEQADVASSVLEFGN 585
QY 522 ---PALSRVSRSIGESGRSADASORAAETIVRDSQTLGDVYSRL-----QVILSL 570
DB 586 FCDKACMTTJTVSNG-----AEVKFASVLLKRGQATATAYTNLKNSSYNGKIKSYV 635

QY 571 MSTIVSNPOANOEIMQKLTASISKAPGYPVAVNSADSLQKFAQLERE---FVDCGR 627
DB 636 VKRYIVDPKSKQONPNGVWMLGIFPDPILGVRA-----SAYIQNKDKDISIFIKNEF 687
QY 628 SLAEQSENAFRKQPAFIQVLVNIASL 654
DB 688 TEYDEGDN-----PIDFDNALLSVASL 709

RESULT 30
MSS_CYPCA
ID MSS_CYPCA STANDARD; PRT; 1935 AA.
AC 090339;
DT 16-OCT-2001 (Rel. 40, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Myosin heavy chain, fast skeletal muscle.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Cyprinus.
OX NCBI_TaxID=7962;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fast muscle;
RX MEDLINE=97352533; PubMed=9208928;
RA Hirayama Y., Watabe S.;
RT "Structural differences in the crossbridge head of temperature-associated myosin subfragment-1 isoforms from carp fast skeletal muscle.";
RT Eur. J. Biochem. 246:380-387(1997).
RN [2]
RP SEQUENCE OF 981-1935 FROM N.A.
RC TISSUE=Fast muscle;
RX MEDLINE=97176447; PubMed=9023993;
RA Imai J., Hirayama Y., Kikuchi K., Kakinuma M., Watabe S.;
RT "cDNA cloning of myosin heavy chain isoforms from carp fast skeletal muscle and their gene expression associated with temperature acclimation.";
RT J. Exp. Biol. 200:27-34(1997).
RN [3]
RP SEQUENCE OF 1387-1528 FROM N.A.
RX MEDLINE=95194396; PubMed=7887920;
RA Watabe S., Imai J., Nakaya M., Hirayama Y., Okamoto Y., Masaki H., Uozumi T., Hiroo I., Aoki T.;
RT "Temperature acclimation induces light meromyosin isoforms with different primary structures in carp fast skeletal muscle.";
RT Biochem. Biophys. Res. Commun. 208:118-125(1995).
RN [4]
RP FUNCTION: MUSCLE CONTRACTION.
CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEATITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED SUBFRAGMENT (S2).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL; D8992; BAA2069.1; -
DR EMBL; D50476; BAA09069.1; -
DR EMBL; D43700; BAA07802.1; -
DR HSSP; P13538; 2MS.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head.1.
DR Pfam; PF00612; IQ.1.
DR Pfam; PF01576; Myosin_tail.1.
DR Pfam; PF02736; Myosin_N.1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head.1.
DR SMART; SM00015; IQ.1.
DR SMART; SM00242; MYSC.1.
DR PROSITE; PSS0096; IQ.1.
DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW ATP-binding; Methylation; Alkylation; Calmodulin-binding;
KW Multigene family.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT DOMAIN 782 811 IQ.
FT DOMAIN 812 839 HINGE.
FT DOMAIN 840 1935 COILED COIL (POTENTIAL).
FT NP_BIND 178 185 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING.
FT DOMAIN 761 775 ACTIN-BINDING.
FT MOD_RES 129 129 METHYLATION (TRT-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1).
FT MOD_RES 709 709 ALKYLATION (SH-2).
SQ SEQUENCE 1935 AA; 22159 MW; 9A1244B67D63C83B CRC64;

```

```

Query Match 5.9%; Score 187.5; DB 1; Length 1935;
Best Local Similarity 22.0%; Pred. No. 1.3;
Matches 158; Conservative 89; Mismatches 265; Indels 205; Gaps 31;

```

```

QY 14 SVSSNOSNMPIINGOIASNSETKESTKASEASPSASSVSSWFLSSAKNALISLSDAT 73
DB 852 AMKENYEKKEDLTAKLAKKELEEKWVS-----LIQEKNDL-QLVTA 894
QY 74 LKNSSEPTD-----SLQLEASTSTSTVTVAAKDYDEAKSNDTAKSGLENNKTLA 125
DB 895 ESENLSDAERCGELTKSKTQLERAKLE--TWERLEDEETIAELTAKRKRLDECESEL 951
QY 126 EYETKADLMAALQDMERLANSDPNNHTE-----VANIKKAL-EAOKDTIDKL 174
DB 952 KRIDDELLETLAVKEKATEKVKNLFEEMASQDESTAKLTKKEKALQEAHQITLDDL 1011
QY 175 NKLVTLQONOKSLTEVLKTTDSADQIPAINSOLEINKNSADQITIKLERONISYEAVLTN 234
DB 1012 Q-----AEEDKVTTLTKAKTKLEQVDDLEGSLEOEK---KLMDLERAKRKLECDLKL 1062
QY 235 AGEVIT-----KASSEAGTK-----LGQALQSTIVDAGDOSQAVALQOONNSPDNTAATK 283
DB 1063 AQSISIMDLENKQSDKEKIKKDFEISQLSKIED--EQLSGAQDK-----IK 1110
QY 284 ELIDAEFTVNELKQHTGLTDSPLVKAKEQISQAOKDOETKPGSDIPYVPGSGSAA 343
DB 1111 EL-----QARIEELEEEIE--AERAARAYEKQADLSRELEI-----SERLE 1152
QY 344 SAGSAAAGALKSSNNSGRISILLDDVDNEMAATALAGFRSMITOFNNPATATKELQAMEA 403
DB 1153 EAGGATATA-----QIEK-----NKKREAFQKMRRLDESTLQHEATATAALRKQEA 1198
QY 404 QLTAMDOVLGADGELPAELQAKDALQALAKO--PSAGGLATAMQOVAFAPAAKVGGSAA 461
DB 1199 DSYA-----EL-----GEQIDINQRYKQKLEKESYKMEITDLTNSMDEVAKAKA----- 1244
QY 462 GFAGTQVMVKOLYKTAFFSTSTSSVAALSDGYSAYKT-----LNSLYSESRGQV 513
DB 1245 -----NLEKKCRT-----LEDLQSLSEIKTKSDENVQLNDM-----NQO 1277

```

```

QY 514 SAISQTPANPALSRVS-----RSGIESQGRSADASQARAETI 550
DB 1278 RARLQTENGEFSRQLEKEKALVSQLTGRKOAYTQOIEELKRIHEEVKAKNALAHAVQSA 1337
QY 551 VRDSQTLGDIYSRLQYDLSIMSTIVS--NQANQ-----EELMQLTIS 592
DB 1338 RHQCDLLREQYEEQEKAKELQGMKANSVAKWMTKRYETDAIQRTTELEEKKKLAQR 1397
QY 593 ISKAPQGYPAVONSADSLQKFAQLERE-----FVYDGER--SLAESQENAFKRPAPF 643
DB 1398 LQDAEE-STEAIVNSKCAKSLKTKQRLQGEVEDLMIDVERANSILA---ANLDKQRNF 1450

RESULT 31
MYSS_RABIT
ID MYSS_RABIT STANDARD; PRT; 1084 AA.
AC P02562;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Myosin heavy chain, skeletal muscle (Fragments).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE OF 1-258.
RA Capony J.-P., Elzinga M.;
RT "The amino acid sequence of A 34,000 dalton fragment from S-2 of
RT myosin.";
RL Biophys. J. 33:148A-148A(1981).
RN [2]
RP SEQUENCE OF 259-428.
RX MEDLINE=85131142; PubMed=3972832;
RA Lu R.C., Wong A.;
RT "The amino acid sequence and stability predictions of the hinge
RT region in myosin subfragment 2.";
RL J. Biol. Chem. 260:3456-3461(1985).
RN [3]
RP SEQUENCE OF 409-1084 FROM N.A.
RX MEDLINE=87304245; PubMed=3305014;
RA Maeda K., Sczakiel G., Wittinghofer A.;
RT "Characterization of cDNA coding for the complete light meromyosin
RT portion of a rabbit fast skeletal muscle myosin heavy chain.";
RL Eur. J. Biochem. 167:97-102(1987).
CC 1- FUNCTION: MUSCLE CONTRACTION.
CC 1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC 1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC 1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC 1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC 1- SIMILARITY: THE PERIODICITIES OF HYDROPHOBIC AND CHARGED RESIDUES,
CC WHICH DICTATE THE ALPHA-HELICAL COILED-COIL STRUCTURE ARE
CC CONSERVED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X05958; CAA29391.1; -
DR PIR; A02985; A02985.
DR PIR; A05280; A05280.
DR PIR; S00084; S00084.

```

KM Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KM ATP-binding; Multigene family.
 FT NON_TER 1 1
 FT DOMAIN <1 >258 ALPHA-HELICAL TAIPIECE (S2).
 FT NON_CONS 258 259
 FT DOMAIN <259 1084 RODLIKE TAIL (S2 AND LMM DOMAINS).
 FT DOMAIN 455 1084 COILED COIL (POTENTIAL).
 FT VARIANT 405 405 L -> V.
 FT VARIANT 408 408 V -> L.
 FT VARIANT 421 421 E -> D.
 FT VARIANT 423 423 S -> G.
 FT VARIANT 426 426 K -> R.
 SQ SEQUENCE 1084 AA; 12548 MW; 229CFD69A6E1F7F0 CRC64;
 Query Match 5.8%; Score 187; DB 1; Length 1084;
 Best Local Similarity 21.4%; Pred. No. 0.69; Indels 174; Gaps 26;
 Matches 145; Conservative 94; Mismatches 265; Indels 174; Gaps 26;
 QY 29 QIASESTKSTKASESPASSSVSWFSSAKNALISLDAIINKNSPTDLS--- 85
 DB 7 EVANKKEEKEKTESLAKAEKRE-----LEEKWALMOEKNDLOLOVQAEADSLADAE 61
 QY 86 -----OLEASTSTSTVTRAAKDYDEAKSNEDTAKSGLENAKTIAEYETKADIM 135
 DB 62 ERQDLIKTKIQLEAKIKEYT---ERADEDEEINAEITAKKRRLDECSSELKDDIDLELT 118
 QY 136 AALQDMERLANSDPSNNHEEYNNI-----KKAL-ENQKOTIDILNKLYTQNON 184
 DB 119 LAKVEKEKATENKYNKLTLEEMAGLDETIAKTKERKALQEHQQLDLDQ-----AEED 173
 QY 185 KSLTEFLKTTSDADQIPAINSOLEINKNSADQIINKLEHONISYEAVLNNAGEVIKASE 244
 DB 174 KNTLTILKTKTKLEQVYDDEGLESEDEK---KIRMDLER-----AKRKLE 214
 QY 245 AGIKGQALQISTVDAGDSQAALVLAQONNSPDNIAATKELIDAET---KYNELKQEH 301
 DB 215 GLTKLAQ--ETSMDEIENDQQLDEKIKLEFMTNLOSKTIEDQALMTNLTQTELEEEIE 272
 QY 302 GLTDPVYKKAEEQISOQKQIOETKPSGSDPIYVPGSSASAGSAAALAKSSNNSGT 361
 DB 273 AERASR--AKAEKQKSDLSRELEET---SRLEBAGA-----TSAQI 310
 QY 362 SLILDDVDNEMAAIALQGRSMIEQFNVNPNPATAKELQAMEAQLTAMSDQLVGADGELPA 421
 DB 311 EM-----NKKREAEFEKRRDLEENTLOHEATPAALAKKHDSVA--EL--GEQID 357
 QY 422 ETQAIKDALAQALKQ--PSADGLATAMGOVAPAAKVGGSAGTACTYOMNKKLYKTA 479
 DB 358 NLQRYKOKLEKEKSEIKMEIDDLA-----GNMETYSKAKGNLEKMCRT-- 400
 QY 480 SSTSSSYAALSDGSAVKTLINSUXE--SRSGVQSAISQTAANPALSR-----S 527
 DB 401 -----LEDOLSEKTKKEEERQRLINELISAQKALHTESGEFSRQDLEKDMVSO 449
 QY 528 VSRSG-----IESQSRASDASQRAEETVRSQTLGDIVYSRLQVLDLSMTY 574
 DB 450 LSRGQAFQTOIEGLKRLQEEETKAKSALAHALQSSRRCDLLREYEEEOEAKAELOA 509
 QY 575 VS--NPQANO-----EETMOKITASISKAPFGYPAYONASDLSQKFAA 616
 DB 510 MSKASEVSOQMTKCTETDIAQRTTELEEAKKRLAQRLQDAEERH-VEAVNSKASCEKTKQ 568
 QY 617 QLERE---FVDEGERSLA 630
 DB 569 RLQNEAREDLMDIVERSNA 586
 RESULT 32
 MYHT_PIG STANDARD; PRT; 1935 AA.
 AC P79293;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, cardiac muscle beta isoform (MyHC-beta).
 GN MYH7.
 OS Sus scrofa (pig).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 CC NCBI_TaxId=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Domestica;
 RA Ko T.L.;
 RL Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
 CC -I- FUNCTION: MUSCLE CONTRACTION.
 CC -I- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -I- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -I- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -I- PFM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -I- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -I- MISCELLANEOUS: THE CARDIAC ALPHA ISOPFORM IS A 'FAST' ATPASE
 CC MYOSIN, WHILE THE BETA ISOPFORM IS A 'SLOW' ATPASE.
 CC -I- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -I- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U75316; AAB37320.1; -
 CC DR HSSP: P08799; IAMD.
 CC DR InterPro: IPR000048; IQ_region.
 CC DR InterPro: IPR004009; Myosin_N.
 CC DR InterPro: IPR002928; Myosin_tail.
 CC DR InterPro: IPR001609; myosin_head.
 CC DR Pfam: PF00063; myosin_head; 1.
 CC DR Pfam: PF00612; IQ; 2.
 CC DR Pfam: PF01576; Myosin_tail; 1.
 CC DR Pfam: PF02736; Myosin_N; 1.
 CC DR PRINTS: PR00193; MYOSINHEAVY.
 CC DR PRODOM: PD000355; myosin_head; 1.
 CC DR SMART: SM00015; IQ; 1.
 CC DR SMART: SM00242; MYSC; 1.
 CC DR PROSITE: PS00096; IQ; 1.
 CC DR Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 CC ATP-binding; Methylation; Alkylation; Multigene family;
 CC Calmodulin-binding.
 CC KW Calmodulin-binding.
 CC FT DOMAIN 1 780 MYOSIN HEAD-LIKE.
 CC FT DOMAIN 781 810 IQ.
 CC FT DOMAIN 839 1935 COILED COIL (POTENTIAL).
 CC FT NP_BIND 178 185 ATP (POTENTIAL).
 CC FT DOMAIN 655 677 ACTIN-BINDING.
 CC FT DOMAIN 757 771 ACTIN-BINDING.
 CC FT MOD_RES 129 129 METHYLATION (THI-1) (POTENTIAL).
 CC FT MOD_RES 695 695 ALKYLATION (SH-1) (POTENTIAL).
 CC FT MOD_RES 705 705 ALKYLATION (SH-2) (POTENTIAL).
 CC SQ SEQUENCE 1935 AA; 223110 MW; 149CDBFD910DBB08 CRC64;
 Query Match 5.8%; Score 187; DB 1; Length 1935;
 Best Local Similarity 19.4%; Pred. No. 1.4; Indels 128; Gaps 24;
 Matches 137; Conservative 134; Mismatches 307; Indels 128; Gaps 24;
 QY 8 MESGPEVSSNQSNMPLINGQIASNSERTKASESPASSSVSWFSSAKNALI 67

```

DB 891 VOAEODNLSDESRCDOLKKNKIOLEAKKEMTERLEDEEENNAELTA----- 938
OY 68 SLRDAIINKNSPTDLSOLEASTSTSYTRVAANDYDEAKSNFPAKSGLEMANLTALEY 127
DB 939 -----KKRVEDBCSELKRDIDDELTL--LAKVEKKAHTENKVNKLTEEMAGLDEI 988
OY 128 ETKMADLMAALODMERLANSPPSNHTEEVNNIKKALEAKDITDKL----- 174
DB 989 IAKLTKREKALQEAHQQA--LDLOLQAEEDKVNLTAKAVKLEQHVDDLAGSLBOEKKVRMD 1047
OY 175 -----NKLVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNSADQII 218
DB 1048 IERAKRKLEGLDKLTQESTMDLENKQQLDERLKKDF--ELNALNARLEDQALCSQQL 1105
OY 219 KDL-----ERONISYEAVLTNAGEVIKASSEAGIKIGQALOSIYDADGSOQAIVLQAL 271
DB 1106 KILKELQAIIELELELELEKERTAKAVKELRSDLSRELEFEISERLEAGAGATSVQI--EMN 1164
OY 272 QNNSPDNIATKELIDAA--ETKYNELKOEHTGLTDSPLVKAEBQISOAKDIOEIKP 328
DB 1165 KKRBAEFQKMRDLDEATLQHEATAALRRKH--ADS--VALGQIDNLQRVKQKLEK 1219
OY 329 SCSDIPIVGPSSASAGASAGALAKSSNNGRISLLDQVNDMAIALQGRSMIE--- 385
DB 1220 EKSEFKELIDVYTS---NMEOIIRAKANLEKMCRTLEQDMEHRSKAEQTSRVNDLTS 1275
OY 386 ---QFVNVPATAKELQAMEA---OLT---AMSDQVYAGDELPEIPEIQAIDALQALK 435
DB 1276 QAKAKIOTENGELSKRDDEKALISQLTRGKLTYYTOLEDKRQLEEVKA--KNALAHALQ 1334
OY 436 --QPSADGLATMGQVAFPAARVGGG--SAGTAGVOMNVK---OLYKTAFFSTSSSYA 488
DB 1335 SARHADLRLREYEEETETKAEQRLVLSKANSEVQAQWRTKETDAIQRELELEAKKLA 1394
OY 489 AALSQYSAYKTLN---SLYESRSGVQSAI-----SOTANPALSRSVS----- 529
DB 1395 ORLODAEEVEEAVNANAKSSLEKTKHRLQNEIEDLMVVERSNAAAALAKKORNFDKILA 1454
OY 530 --RSGLIESGRSADSORAETIVRDSQTLGDVYSRLQVLSMSTIVSNPQANOBEIMO 587
DB 1455 EMKQKKEESQSELESQKARSLSTELFKLKNYE--ESLEHLETKRREN--KNLOEETS 1510
OY 588 KITASISKAPQFGYPAVONSADSLQFAAQLEREFEVGGERSLAESQ 633
DB 1511 DLTEQGGSSGK-----TIHELEKVRQLERAEKLEQSALEESE 1548

RESULT 33
MYB1_LYCES STANDARD; PRT; 697 AA.
AC P93203:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAR binding filament-like protein 1.
GN MYB1.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv VENT Cherry; TISSUE=fruit;
RX MEDLINE=97112038; PubMed=8953774;
RA Meier I., Phelan T., Grussem W., Spiker S., Schneider D.;
RT "MYB1, a novel plant filament-like protein with affinity for matrix
RT attachment region DNA."
RL Plant Cell 8:2105-2115(1996).
CC -I- FUNCTION: BINDS DNA. INTERACTS WITH CHROMATIN VIA MATRIX
CC ATTACHMENT REGIONS (MARS). LIKELY TO PARTICIPATE IN NUCLEAR
CC ARCHITECTURE BY CONNECTING CHROMATIN WITH THE NUCLEAR MATRIX AND

```

```

CC CC POTENTIALLY WITH THE NUCLEAR ENVELOPE.
CC -I- SUBCELLULAR LOCATION: NUCLEAR MATRIX.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC CC
DR EMBL: Y07861; CAB69181.1; -.
KW Nuclear protein; DNA-binding; Coiled coil.
FT DOMAIN 125 681 COILED COIL (POTENTIAL).
FT DOMAIN 10 15 POLY-SER.
SQ SEQUENCE 697 AA; 79516 MW; 700A56D68D6A7E49 CRC64;

Query Match 5.8%; Score 186; DB 1; Length 697;
Best Local Similarity 19.7%; Pred. No. 0.44;
Matches 98; Conservative 84; Mismatches 151; Indels 164; Gaps 19;

OY 29 QIASNETKESYKASBPASSSVSWGFLSSAKNALISL-----RDAILINKNS---SPT 81
DB 246 QVLEEKITLITTEIKDKVSLSNSTSKLAKESYVNSLSDMYQOSODOLMNLSTELKEL 305
OY 82 DLSOLEASTSTSYVT-----RVAAKYDEAKSNFDTAKSGLE 119
DB 306 DETQKRELELEKCVSEDNLANQLNSLLERDESKELHAIQKEYSEFKNSD---EKVA 362
OY 120 NAKTAEYETKMAADLMAALQDMERLANSPPNNH-----TEEVNNIKKALEAKQDITDK 173
DB 363 SDATLGEQKRLHQ---LEQLGTALEASAKNEVLIADLTKEKELRMVALELDNVK 418
OY 174 LNK--LVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNSADQIIKDERONISYE-- 229
DB 419 LKQEIFVYTESLENSRSEVSDITVQLELRDLSKIE---REVSKLQMLEETRSLQGN 475
OY 230 -----AVLTNAGEVIRKASSEAGIKIGQALQISVDAGDQSOAAVLAQOQNSPD 277
DB 476 IDETKHSELLAEELTJTKELKTNE-----EMHTMSD 509
OY 278 NIAATKELIDAEKTV-----NELKOEHTGLTDSPLVKAEBQISOAKDI--- 323
DB 510 ELVAIVSENDSLOTLELVNYYKKRREHRELKQEK---IVRTLEELKLELSQITRE 563
OY 324 QETKPSGSDIPIVGPSSASAGASAGALAKSSNNGRISLLDQVNDMAIALA----- 376
DB 564 KELRKSLED-----ELEKATES-----LDEINRNVLALELELAT 599
OY 377 -----LQGRF---SMIDQFVNPNPA-----TAKEQAMEAQLTMSDOLV 413
DB 600 SRNSLEDEREVHROGSVBOKISOEAQENLEDAHSLVVKLKERESLEKRAKKLEDEMA 659
OY 414 GADGE---LPAELQAIK 427
DB 660 AAKGELLRLRSQINSVK 676

RESULT 34
MYHB_HUMAN STANDARD; PRT; 1972 AA.
AC P35749; O00396; P78422; O94944;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, smooth muscle isoform (SMMC).
GN MYH1 OR KIA0866.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE-99425270; PubMed-10493829;
 RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
 RA Fuhman J., Mason T., Crosby M.L., Barnstead M., Crohn L.,
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
 RA Richter E.E., Harris P.C., Venter J.C., Adams M.D.,
 RT "Genome duplications and other features in 12 Mb of DNA sequence from
 RT human chromosome 16p and 16q.";
 RL Genomics 60:295-308(1999).
 RN [2]
 RP SEQUENCE OF 1-1266 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE-99156230; PubMed-10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364(1998).
 RN [3]
 RP SEQUENCE OF 885-1972 FROM N.A.
 RX MEDLINE-93263189; PubMed-7684189;
 RA Matsuo R., Yoshida M.C., Fututani Y., Imanura S., Kanda N.,
 RA Yanagisawa M., Masaki T., Takao A.;
 RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal
 RT region 16q12.";
 RL Am. J. Med. Genet. 46:61-67(1993).
 RN [4]
 RP SEQUENCE OF 1093-1972 FROM N.A.
 RC TISSUE=Hippocampus;
 RA Okajima K.;
 RL Submitted (NOV-1992) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- TISSUE CELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: SMOOTH MUSCLE; EXPRESSED IN THE UMBILICAL
 CC ARTERY, BLADDER, ESOPHAGUS AND TRACHEA.
 CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PMW: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- DISEASE: A chromosomal rearrangement, known as pericentric
 CC inversion inv(16)(p13q22), produces a fusion protein that consists
 CC of the 16S N-terminal residues of CBF-beta (PEP82) with the tail
 CC region of MYH11. This rearrangement is associated with acute
 CC myeloid leukemia of M4EO subtype.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEMOROUSIN (LMW) AND 1 HEAVY MEMOROUSIN (HMW). IT CAN LATER BE
 CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isdb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF001548; AAC31665.1; -;
 DR EMBL: U91323; AAC35212.1; -;
 DR EMBL: AB020673; BAAY4889.1; -;
 DR EMBL: D10667; -; NOT_ANNOTATED_CDS.
 DR EMBL: X69292; CAA49154.1; -;
 DR HSSP: P10587; 1BR2.
 DR GENE: HGNC:7569; MYH11.
 DR MIM: 160745; -;
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.

DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head.
 DR Pfam: PF00612; IQ: 1.
 DR Pfam: PF01576; Myosin_tail: 1.
 DR Pfam: PF02736; Myosin_N: 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRODOM: PD00035; myosin_head: 1.
 DR SMART: SM00015; IQ: 2.
 DR SMART: SM00242; MYSC: 1.
 DR ProSITE: PS50096; IQ: 1.
 KW Myosin: Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family; Proto-oncogene; Chromosomal translocation.
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
 FT DOMAIN 786 815 IQ.
 FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
 FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 762 776 ACTIN-BINDING (BY SIMILARITY).
 FT MOD_RES 129 129 METHYLATION (TR1-) (POTENTIAL).
 FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
 FT CONFLICT 887 889 EEL -> NSE (IN REF. 3).
 FT CONFLICT 1263 1266 ELQS -> TLST (IN REF. 2).
 FT CONFLICT 1558 1558 T -> S (IN REF. 3).
 FT CONFLICT 1610 1611 RO -> NE (IN REF. 3).
 FT CONFLICT 1786 1786 A -> S (IN REF. 4).
 FT CONFLICT 1958 1958 T -> L (IN REF. 3).
 SQ SEQUENCE 1972 AA; 227338 MW; 67665BBAECEL277 CRC64;
 Query Match 5.8%; Score 186; DB 1; Length 1972;
 Best Local Similarity 21.4%; Pred. No. 1.6;
 Matches 141; Conservative 108; Mismatches 270; Indels 140; Gaps 29;
 QY 111 PPTASGLENATLTAEYFKMDLMAALQDMERLAN--SDPENNNH---TEEVNNTKKALE 165
 DB 838 FTKVAPLLOVTRQOEEMQKDELOKTKRQKAELKELBOKKSQLEENKLLQEOQL 897
 QY 166 AOKDTIDKLNKT-VTLQONKSLTEVLYKTTDS-ADQIPAINSOLEINKNSADQIKDLER 223
 DB 898 ATELYAEAEEMKRVRLAAKKQELFEIHEMARELEEDRGQQLQAEKKMAQOMDLDEE 957
 QY 224 QNISTEAVLTNAGEVYKASSEGIKGLQALQSIDVADGOS-----QAAYLQAOQNNSPUN 278
 DB 958 Q-LEEEEAARQRLQLEKYVAEAKIK--KLEDELIVMDQNNKLSKERKLEERISDLTTN 1014
 QY 279 IATKE-----LIDAAETKYNEL-----KQEHGLTDSPYVKRAE-----EQISQA 319
 DB 1015 LAEEERKANLTKLNKHSMTSELEVRLLKKEKSKQELKLRKLEGDSPHEQIDAL 1074
 QY 320 QKDIOEKPSGSDIPYVSGSAASGSA--GALKSSNN--GRISLLLDVDD 369
 DB 1075 QAQIAELKNG-----LAKKEBELQAALARLDEIAQKNNAKLTRELEGHISDLQEDPD 1128
 QY 370 NEMAA-----IALQGFRTNEDQFNVPNPAKELQAM----- 401
 DB 1129 SERAARNKAERKQKRDJGEELEALKTELED-TLDSTATOELAKRBOEVTYVKALDEET 1187
 QY 402 ---EQLTMSQDLVGADELPABEIOATKDALA-----QALKQPSAD--GLATMNG--- 447
 DB 1188 KSHENAVQEMRKHAQAVVELTQLEQFRAKAAANDKKNQITLEKKNADLAGELRYLGQAK 1247
 QY 448 -----QVAFPAAKVGGSGAGTAVQVM--VKQLYKTAFSSSTSSSYAALSD 493
 DB 1248 QEVHKKKKLEAVOELQKSCDGERARA---ELNDKVKHKLNEVESYVGMLEAE--GK 1302
 QY 494 GYSAYKTLNLSYSEBSRGVQSAISOTANPALRSYRSRISQSGS-----AASQRAET 549
 DB 1303 AIKLKDVASLSSLOD--TOELLQETROKLVNSTRKLRLEERNSLQDQLEEMERAKON 1361

```

OY 550 IYRDSOT----LGDVYSRLQVLDLSMTIVSNPQANOEIMOKLTAISKAPQFCYPANQ 605
DB 1362 LERHISTLNIQSDSKKKLDPASTVEALEEGKKRQKEI-ENL100-----YEKA 1412
OY 606 NSADSLQKFAAOLEREFE-----VDGERSLAESENAFROPAFIOQVLVNIASLPSGX 658
DB 1413 AAYDKLEKTKRNLQQLDLDLVLDNQOLVSNLEKKOKK----FQDLAEKKNISKY 1467

RESULT 35
AKA9_RABIT
ID AKA9_RABIT STANDARD: PRT: 1087 AA.
AC Q28628:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRA9) (A-kinase anchor protein 120 kDa) (AKAP 120) (Fragment).
GN AKAP9 OR AKAP120.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Gastric parietal cell;
RX MEDLINE=97220389; PubMed=9148752;
RA Dransfield D.T., Yeh J.L., Bradford A.J., Goldenberg J.R.;
RT "Identification and characterization of a novel A-kinase-anchoring
RT protein (AKAP120) from rabbit gastric parietal cells."
RL Biochem. J. 322:801-808(1997).
CC -I- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
CC A. MAY BE A SCAFFOLDING PROTEIN.
CC -I- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN GASTRIC PARIETAL CELLS.
CC -I- DOMAIN: RTI BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
CC COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: U26360; AAC35413.1; ALT_INIT.
CC KW Coiled coil.
FT NON_TER 1 1
FT DOMAIN 559 572 PRA-RTI SUBUNIT BINDING DOMAIN.
FT DOMAIN 5 461 COILED COIL (POTENTIAL).
FT DOMAIN 614 773 COILED COIL (POTENTIAL).
FT NON_TER 1087 1087
SQ SEQUENCE 1087 AA; 124756 MW; 9D916BEOCA9FP02 CRC64;

Query Match 5.8%; Score 185.5; DB 1; Length 1087;
Best Local Similarity 22.3%; Pred. No. 0.8;
Matches 139; Conservative 98; Mismatches 242; Indels 145; Gaps 27;

```

```

DB 335 EEKNELIRLEAOIECL-----MSDQERYKKNREBEITQNLNEVIEKLOELANIDOKTS 388
OY 336 VGPSSGASAGSAGALKSSNNSGRISLLDDVDNEMAAIALQGRSMIEGFNNVPATA 395
DB 389 VDPS-----SLSEADSLKHQ-----LDKVIKALKLEHQ-----VETNEEMAYTK 430
OY 396 KELQAMEAOLTRMSDLQ--VGADGELPAETQAI-----KDA 429
DB 431 NVLKEFNFKMNOITQELCSLKREREMERIQSVPEKSVNMSVGLSKDKPPEMDLIPTEDA 490
OY 430 LAQALROP-----SADGLATAMQVAFPAAKVGSAGTAGTVQMVVKOLYK-- 476
DB 491 LQOLETQIQLRSESESKKSLSLSEKLIQL-----ESTVSTDELLTQCYKQI 539
OY 477 -----TAFSSSTSSSSVYAAALSDGYSAYKTINSLYSESR-SGVQSAISQTN-PAL 524
DB 540 QDMREGRETEMLQRIKYSILQVLEE-----KVAALVSQVLEAVQYVKLADKPAY 594
OY 525 SRSVSRSGIESQGRSADASQRAETIVRDSQTLGDIYSHLQ-VLDSLMTIVSNQANDE 583
DB 595 SSDPATEVPG-----LSQLAGNTEMSDVSALTRISIELSQVLEMHSLIS--EKDQV 646
OY 584 EIMOKLTASISKAPQFCYPANQNSADSLQKFAAOLERE---FVDEGRSLAES----- 632
DB 647 EIAEKNALEKEKKLQELQKLVQDSE-----TKQREKRSKRLHGDGLVLESTSEESGV 700
OY 633 --QENAFKROPAFIOQVLVNIASL 654
DB 701 FQELALRASAAPKELANYKEL 724

RESULT 36
SMC_MYCTU
ID SMC_MYCTU STANDARD: PRT: 1205 AA.
AC Q10970:
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Chromosome partition protein smc.
GN SMC OR RV2922C OR MT2990 OR MTCY338.11C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA Cobbe N., Heck M.M.S.;
RT "Phylogenetic analysis of SMC proteins."
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagsels K., Kirog A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Bartell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;

```


RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN CHROMOSOME STRUCTURE AND
 CC PARTITIONING. ESSENTIAL FOR CHROMOSOME PARTITION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ141609; CAC93884.1; -
 DR EMBL; Z74697; CAA98982.1; ALT_INIT.
 DR EMBL; AE007121; AAK47317.1; -
 DR TIGR; MT2990; -
 DR TubercuList; RV2922c; -
 DR InterPro: IPR003439; ABC_transport.
 DR InterPro: IPR003405; SMC_C.
 DR InterPro: IPR003395; SMC_N.
 DR Pfam; PF02463; SMC_N; 1.
 DR Pfam; PF02483; SMC_C; 1.
 DR ATP-binding; Coiled coil; Complete proteome.
 DR NP_BIND 31 38 ATP (POTENTIAL).
 FT DOMAIN 167 289 COILED COIL (POTENTIAL).
 FT DOMAIN 330 499 COILED COIL (POTENTIAL).
 FT DOMAIN 659 842 COILED COIL (POTENTIAL).
 FT DOMAIN 899 929 COILED COIL (POTENTIAL).
 FT DOMAIN 979 1038 COILED COIL (POTENTIAL).
 SO SEQUENCE 1205 AA; 130637 MW; A3B2A813B58EAC33 CRC64;

Query Match 5.8%; Score 185.5; DB 1; Length 1205;
 Best Local Similarity 21.7%; Pred. No. 0.91;
 Matches 149; Conservative 103; Mismatches 290; Indels 145; Gaps 30;

QY 30 IASNSTKSTKASEKSPSSVSSMSFLSSAKKALISLRDLILKNSPTDLSQLRA 89
 DB 264 VASEELAAHSAVAELSTPHESTQHTWFGIS---ALAEKVDVTVIAEERAHNLD--- 315
 QY 90 STSTSTVTVAADKYDEAKS---NFDTKASGLENAKTIAEYETKMDLMAA---LQDMER 143
 DB 316 -----IEPVAVSDTPRKRPELELAQOVAAEQOLLLEDAARLRDAARELDRER 369
 QY 144 LANSDFSNHTEEVNNIKKALEKQDTIDKLKLVTLQNNQNSLVEVLTSTDSADQIPAI 203
 DB 370 RA-AEADRAH-----LAAVREADRREGLARLAGOVETMRARVESID--ESVARL 416
 QY 204 NSOLEINKKSADQIILKDLERONISYFVLTNMGEEVTKASSEAGIKIGQALQSYDA---G 260
 DB 417 SEETIEAAMRAQO-----TRAFFETVQGRIGEL---DQGEVGLDEHNERTVAALRLA 465
 QY 261 DOSQAALVLAQONNSPDNIATKELIDAEETKYNEL-----KOEHTGLDPSLVKKA 312
 DB 466 DE-RVAELQSAERAAERQVASLRARIDALAVGLQKDGAAMLAHNNSGAGLGS----- 518
 QY 313 EEOISQAQKDIQEKPSGSDPIVPGSGAA-----SAGSAGALKSSNNSGRISL 363
 DB 519 ---IAQLVK-----VRSGEYEAALAAALGPAADLAVDGLTAGSVAVALKQA-DGSAVL 569
 QY 364 LDD-----VDNEMAIALQGRSMIEQFNVNPNATKELQAMEQOLTRMSQLYG 414
 DB 570 VLSMDWPAPOASASGEMLPSSGQWMLDVE---SPQLVGMAMTMSGLVAVND-LTE 624
 QY 415 ADG--ELPAEIQAIKDALAQALAKOPADGLATFATMGQVAPAAKVGSGSAGTGYOMANK 472
 DB 625 ANGLVIRELRARV-----TVDDGLVAGAGVNS-----GGSDDKSLTILEV-YS 665
 QY 473 QLYKTAESSTSSSYAALSDGYSAVTKTNSIYSESRSQVOSAI--SQTANPLASRSVR 530
 DB 666 EIDKARSELAAAEALAAQALNNAALAGALTQSANQADAAEDALAAALNNSDTAISAMVQQLR 725

QY 531 SGIESQGRSAD-----ASQRAETIVDSQTLGIVY-----SRLQVILDSIMST 574
 DB 726 LGEAAHAAEEENRRLQQTBOEPAVA--TOTLDVIOLETQLKRAQTOYQVQAQPIDRA 784
 QY 575 VS--NPQANQEIIMQITATASISKAPOGYPAYONSADSLQKRAOLREBVDGERSIAES 632
 DB 785 ISAAADRARGVEARLAV---RTAEERANAVRGRADSLRRAA-----AEREARVRAQ 835
 QY 633 QENAFKQPAFIQOVLYVNIASLPSGTL 659
 DB 836 QARAARLHAAVAAVAVDCGRLLAGRL 862

RESULT 37
 REST_HUMAN
 ID REST_HUMAN STANDARD; PRT; 1427 AA.
 AC P30622;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Reslin (Cytoplasmic linker protein-170 alpha-2) (CLIP-170) (Reed-
 DE Sternberg Intermediate filament associated protein).
 GN RSN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Peripheral blood monocytes;
 RX MEDLINE-92289675; PubMed-1600942;
 RA Bilbe G., Delabie J., Brueggen J., Richener H., Asselbergs F.A.M.,
 RA Cerletti N., Sorg C., Olink K., Jarczyk L., Wiesenanger W.,
 RA de Wolf-Peeters C., Shipman R.;
 RT "Reslin: a novel intermediate filament-associated protein highly
 RT expressed in the Reed-Sternberg cells of Hodgkin's disease.";
 RL EMBO J. 11:2103-2113(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92405160; PubMed-1356075;
 RA Pierre P., Scheel J., Rickard J.E., Kreis T.E.;
 RT "CLIP-170 links endocytic vesicles to microtubules.";
 RL Cell 70:887-900(1992).
 CC -!- FUNCTION: SEEMS TO BE A INTERMEDIATE FILAMENT ASSOCIATED PROTEIN
 CC THAT LINKS ENDOCYTIC VESICLES TO MICROTUBULES.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC, ASSOCIATED WITH THE
 CC CYTOSKELETON.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THE REED-STERNBURG CELLS
 CC OF HODGKIN'S DISEASE.
 CC -!- SIMILARITY: CONTAINS 2 CAP-GLY DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X64838; CAA46050.1; -
 DR EMBL; M97501; AAA35693.1; -
 DR PIR; S22695; S22695.
 DR Genew; HGNC:10461; RSN.
 DR MIM; 179838; -
 DR InterPro: IPR000938; CAP_GLY.
 DR InterPro: IPR001878; Znf_CCHC.
 DR Pfam; PF01302; CAP_GLY; 2.
 DR SMART; SM00343; Znf_C2HC; 1.
 DR PROSITE; PS00845; CAP_GLY_1; 2.
 DR PROSITE; PS0245; CAP_GLY_2; 2.

KW Cytoskeleton; Microtubules; Coiled coil; Repeat; Alternative splicing.
 FT DOMAIN 78 120
 FT DOMAIN 143 204 SER-RICH.
 FT DOMAIN 232 274 CAP-Gly 2.
 FT DOMAIN 304 331
 FT DOMAIN 350 342 SER-RICH.
 FT DOMAIN 1408 1421 COILED COIL (POTENTIAL).
 FT VARSPLIC 457 491 MISSING (IN SHORT ISOPFORM).
 FT CONFLICT 1069 1069 D -> E (IN REF. 2).
 SQ SEQUENCE 1427 AA; 160989 MW; 0A4F16DD94254E8 CRC64;
 Query Match 5.8%; Score 185.5; DB 1; Length 1427;
 Best Local Similarity 20.5%; Pred. No. 1.1;
 Matches 143; Conservative 116; Mismatches 291; Indels 147; Gaps 28;
 QY 5 HHMSEGPSVSSNOSSMPITINGOIASNSEKESKASEAS--PSASSSSMSRLSSA 62
 DB 276 HAYTKIGFSTTPAKAKANAVRVNATTSASLKRSASLSSMSVASSVS---RPS 331
 QY 63 KNALISLRDALINKNSPTDSLQLEASTSTSTVTVRAAKDYDEAKSNEDTAKSG-LENA 121
 DB 332 RIGLLETSSRYARKISGTTALQALKEKQOHTEQLEAERDLERAVAKATSHVGEIEQ 391
 QY 122 KTLA-----EYETKADLMALQDMERLANSPPSNHTFEVNNIKKALEAQDTTD 172
 DB 392 LALARDGHQVLELAKKDQLTMTWEADR-----EKVELLNO-LEEEKRYVE 439
 QY 173 KINKLVTLLNONKSLTEVTKTSDAQPAPAINSLQLEINNKNSADQIKOLERQISEAVL 232
 DB 440 DLQFRFEESTYKGDLEFQTKLEHA-RIKELESLLFEKTKADKIDRELEDTIV---ATV 495
 QY 233 TNAGEVIKASSSEAGIK-----LGQALQSTVDAGD-----QSOAAVLQAOQNN 275
 DB 496 SEKSRIMELEKDLALRVOEVAELRRLESINKPAGDVMSLILQETISSIQEKLVTRTDH 555
 QY 276 PNIIATKELIDAAE-----TKVNE---LKQEHGLDSDPLV---KK 311
 DB 556 QREITSLKEHFGARETHQKEIKALYTATEKLSKENEKSKLEHANKENSVDIALMKRSK 615
 QY 312 AEEQISOAQDIOEIKRPSGDIPIVGPSSGASAGASAGAKSSNNSGRISLLDDVDV 371
 DB 616 LERAINSHQOAMELKVST-----KLGITTAEPH-ELKQIETKML-----DYQHE 662
 QY 372 MAALALQGRSMIEQFNVNPNATAKELQAMEAQLTAMSDQLVAGDELPAEIQAIKDATA 431
 DB 663 I-----ENLONQODSEERAAHAKEMEARAKIMKVIK-----EKENSLAIRSKLD 707
 QY 432 QALKOPSDGLATAMGOVAFPAAKVGGSGAGTGYOMNKOLYKTAFFSTSS----- 484
 DB 708 KAADQHLVE-MEDTLNKLQDAELKV-----KELEVLQAKCNEQTKVIIDNFTSQLKATEEK 761
 QY 485 ----SSYAAALSDGYAYKTLNLSYESRSYGVS-AISQTANPALRSRVSRSIESQGRS 539
 DB 762 LLDLDLAKRASSSGKEMMKLRQOLEAKEQIKHLEIKKAESSKASSSTR---ELQGRE 818
 QY 540 -----ADASQRAELTYVDQSOTLGDVYSR-----LQVLDLSMTSTVSNPQANQEE 564
 DB 819 LKLTNLQENLSEVSO-VKETLEKEQLILKEKFAEASEEAVSVRSQETV--NKLHQKEE 875
 QY 585 IMOKLTASTISKAPQPGYPAVQNSADSLQKFAADLERE 621
 DB 876 QFNMLSSDLEKLR-----ENLADMEAKFRREKDERE 905
 RESULT 38
 MYSC_CAEEL
 ID MYSC_CAEEL STANDARD; PRT; 1947 AA.
 AC P12845;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain C (MRC C).
 GN MYO-2.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=89178677; PubMed=2926820;
 RA Dibb N.J., Maruyama I.N., Krause M., Karn J.;
 RT "Sequence analysis of the complete *Caenorhabditis elegans* myosin
 heavy chain gene family";
 RL J. Mol. Biol. 205:603-613(1989).
 RN [2]
 RP SEQUENCE OF 1-21; 112-371 AND 1501-1772 FROM N.A.
 RC STRAIN=Bristol N2;
 RX MEDLINE=85201409; PubMed=3888374;
 RA Karn J., Dibb N.J., Miller D.M.;
 RT "Cloning nematode myosin genes";
 RL Cell Muscle Motil. 6:185-237(1985).
 CC -1- FUNCTION: MUSCLE CONTRACTION.
 CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
 AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN THE PHARYNGEAL MUSCLE.
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
 ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
 CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
 SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 SUBFRAGMENT (S2).
 CC -1- MISCELLANEOUS: THERE ARE FOUR DIFFERENT MYOSIN HEAVY CHAINS IN
 C. ELEGANS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sb.ch/announce/>
 or send an email to license@isb-sb.ch).
 CC EMBL; X08066; CAA30855.1; -
 DR EMBL; M37233; AAA28121.1; -
 DR EMBL; M37235; AAA28122.1; -
 DR EMBL; M37236; AAA28123.1; -
 DR PIR; S05697; S05697.
 DR HSSP; P10587; 1BR2.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00242; MSc; 1.
 KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW ATP-binding; Methylation; Alkylation; Multigene family;
 FT DOMAIN 1 855 MYOSIN HEAD-LIKE.
 FT DOMAIN 856 1947 RODLIKE TAIL (S2 AND LM DOMAINS).
 FT NP_BIND 174 181 ATP (BY SIMILARITY).
 FT DOMAIN 668 690 ACTIN-BINDING.
 FT DOMAIN 773 787 METHYLATION (TR-1) (POTENTIAL).
 FT MOD_RES 125 125 ALKYLATION (SH-1).
 FT MOD_RES 708 708 ALKYLATION (SH-2).
 FT MOD_RES 718 718 ALKYLATION (SH-2).
 FT CONFLICT 132 132 E -> D (IN REF. 2).

Db 560 TTTTATTANNSTRAATVITDPTLDPDNDNSPND-----NAKSTTYGSSSTGASL 611
 QY 346 GS---AAGALKSSNNSGRISLLDDVD-NEMAAIALOGFRSMTEQFVNPNPAKELQAM 401
 Db 612 DLRRTTTSISVSNMTQVCTSESDYSDSPSPAI-----STATTEENLI 658
 QY 402 EADLTMSDVLVADGLPRAEIOAKKALA-----QALKPSADGLATAMQVAFPAAKV- 456
 Db 659 TMTTAS---CSTDSNFPSPSAASSTDETAFTRTISTSCSTLNGASQTSLELTSPKTN 714
 QY 457 -----GGGSACTAGTVQNNVQLYKTAFFS-----TSSSYAAALSDGYS 497
 Db 715 TVVPASSFPPTTTTCLEN-----DDTAFSSSYTEVNATITINPGTSSLADDFATSEKPN 770
 QY 498 YKTLNLSYSESRSGVOSAIQT-----ANPLSRVSRSGIESGSRSDASQRAAETIVR 552
 Db 771 PTVSVKSTSNEGSTSTTYYQQTVAITLYAKRSPSTSLGARTTGSNGRSTTSQDDGSAM--- 827
 QY 553 DSQGLGVSRQ 565
 Db 828 -HOPTSISYQOLK 839

RESULT 40

MYH4_HUMAN

ID MYH4_HUMAN STANDARD; PRT; 1939 AA.

AC Q9Y623;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Myosin heavy chain, skeletal muscle, fetal (Myosin heavy chain 11b)

DE (MYHC-11b).

GN MYH4.

OS Homo sapiens (Human).

OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=skeletal muscle;

RX MEDLINE=99318869; Pubmed=10388558;

RA Weiss A., Schiaffino S., Ielmand L.A.;

RT Comparative sequence analysis of the complete human sarcomeric myosin

RT heavy chain family: implications for functional diversity.;

RL J. Mol. Biol. 290:61-75(1999).

CC -1- FUNCTION: MUSCLE CONTRACTION.

CC -1- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2

CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)

CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).

CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.

CC -1- DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING

CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,

CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC -1- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY

CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.

CC -1- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT

CC MEROMYOSIN (LMW) AND 1 HEAVY MEROMYOSIN (HMW). IT CAN LATER BE

CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED

CC SUBFRAGMENT (S2).

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC at the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: AF111783; AAD29949.1; .

CC DR HSSP: P13538; 2MTS.

CC DR Genew: HGNC:7574; MYH4.

DR MIM; 160742; -
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR004009; Myosin_N.
 DR InterPro; IPR002928; Myosin_tail.
 DR InterPro; IPR001609; myosin_head.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00612; IQ; 2.
 DR Pfam; PF01576; Myosin_tail; 1.
 DR Pfam; PF02736; Myosin_N; 1.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR PRODOM; PD000355; myosin_head; 1.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00242; MSC; 1.
 DR PROSITE; PS00936; IQ; 1.
 DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
 KW Multigene family.
 FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
 FT DOMAIN 785 814 IQ.
 FT DOMAIN 843 1939 COILED COIL (POTENTIAL).
 FT NP_BIND 179 186 ATP (POTENTIAL).
 SQ SEQUENCE 1939 AA; 223012 MW; 40B1AD1D77A47DE CRC64;

Query Match

Best Local Similarity 19.9%; Score 184; DB 1; Length 1939;

Matches 113; Conservative 104; Mismatches 228; Indels 122; Gaps 22;

QY 102 KDVEAKSNFDTAKSGLENKTLA---EYETKMAADLMAALQDMERLANSDPNNHTEVNV 158
 Db 851 KEANAKKEEKEKRE--ELAKTEAKRELEKAVTLM-----QENK 889
 QY 159 NIKKALEAQKDTI---DKLNKLVTLQNNKSLTEVLKTTDSADQIPAINSOLEINKNSA 214
 Db 890 DLDQVQAFADALDAEERCQDLIKTKIQLEA--KIEVTERADEDEEIVAEITAKRKRL 947
 QY 215 DQITKDLERNITSEAVLTNAGEVITKASPAKIK-----LQALQSLVADGDSQA 266
 Db 948 EDECSLKIDIDLETLKAV-EKEKHATENKVKNLTEENAGLDETAKLT---REKKA 1002
 QY 267 VLOAQNNSPDNIAATKELIDAAETKYNELKQEHGTGSDPLVKKAEQISQAKDIOET 326
 Db 1003 LQEAHQOTLDD-----LQMEDKAVNTLTAKT-----KLEQVDDLESGLSLOC 1045
 QY 327 KPSGSDPIVPGSASASAGSAGALKSSNNSGRISLLDDVDNEMAAIALOGFRSMTEQ 386
 Db 1046 KKLQMDL-----ERAKRKLQGLDLKLAQESTM-----DFENDKQQLNEKLRKEFEW 1091
 QY 387 FVNPNPAKELQAMEKQLTAMSDQVAGDELPAETQAIKKDLAQLKQPSADGLATAM 446
 Db 1092 SNLD--GKIEDQALAMQLOKKIKELQARIELEEEIEAERASRAKAEKORS--DLSREL 1147
 QY 447 GOVAFAAKVGGSAGTAVQNNVQLYKTAFFSSTSSSYAAALSDGYSAYTUNLSVY 506
 Db 1148 EEISERLEEGG--ATSAQIELNKR-----EAERQ 1176
 QY 507 ESRSGV-QSAISQTPANPALSRSVSRSGIESGSRSDASQRAAETIVRSDQGLGVSRQ 565
 Db 1177 KMRDLEESTLQHEATPAALRKKHADSVAELQIDSLQVKKLEKEKSEL-----KME 1231
 QY 566 VLD--SLMSTIVSNPQANQEIIMOKLTASISKAPQGYPAVQNSASLSLQFAALQEREFV 623
 Db 1232 INPLASNMEF-VSKARNFEMKCRITLEDQSL-KTKEERQRLINELSAOKARLHTESG 1289
 QY 624 DGERSLAESQ---ENAFKQPAFIQOV 647
 Db 1290 EFSRQLEKDMVYQSLSRGKQAFITQOI 1316

Search completed: March 12, 2003, 12:10:27

Job time : 74 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 12, 2003, 11:23:47 ; Search time 40 Seconds

(without alignments) 1566.218 Million cell updates/sec

Title: US-10-007-693-139

Perfect score: 3204

Sequence: 1 MHHHHHMHSESPESVSSNQ.....PAFIQVLTNIAISLFGSLYS 660

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2943	91.9	647	2	G71490
2	894	27.9	651	2	D72042
3	894	27.9	651	2	E86381
4	453	14.1	715	2	I40729
5	268	8.4	2271	2	F90073
6	261	8.1	6713	2	B89921
7	235	7.3	2285	2	T12796
8	234	7.3	1063	2	D86731
9	231	7.2	1365	2	T30822
10	226.5	7.1	1269	2	F84730
11	225.5	7.0	1302	1	JC6009
12	224.5	7.0	1790	2	S67593
13	219.5	6.9	2139	2	T18296
14	218.5	6.8	2570	2	T17451
15	218	6.8	4776	2	E35206
16	216.5	6.7	1072	2	A68827
17	214	6.6	2186	2	H89960
18	213	6.6	1690	2	T13030
19	212	6.6	1822	2	S33441
20	211.5	6.6	2481	2	D50011
21	210.5	6.6	1091	2	T34107
22	210	6.5	1104	2	T22976
23	208.5	6.5	1133	2	T22976
24	208	6.5	1306	2	T33318
25	208	6.5	1475	2	C89921
26	208	6.5	3890	2	C82206
27	207.5	6.5	641	1	B43402
28	207.5	6.5	641	1	B43402
29	205.5	6.4	2007	1	B43402

30	204	6.4	1051	2	T18351
31	202.5	6.3	1963	1	MMKW
32	202	6.3	1940	2	A59287
33	201	6.3	2116	2	A26835
34	200	6.2	971	2	B90835
35	200	6.2	973	2	C85693
36	200	6.2	1992	1	S02771
37	200	6.2	2052	2	C97038
38	199.5	6.2	1381	1	S45781
39	199.5	6.2	1957	2	T38077
40	199.5	6.2	2253	2	T30336
41	199	6.2	918	2	C96829
42	199	6.2	1509	1	A27224
43	199	6.2	1940	1	A24922
44	198	6.2	1127	2	T28317
45	197.5	6.2	890	2	S22452

ALIGNMENTS

RESULT 1	G71490	hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UM3/Cx)
N:Alternate names:	chlpn 76kda homolog CT622	
C:Species:	Chlamydia trachomatis	
C:Date:	13-Sep-1998	#sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession:	G71490	
R:Stephens, R.S.; Kaiman, S.; Iammell, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche	Science 282, 754-759, 1998	
A:Title:	Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t	
A:Reference number:	A71570; MUID:9900809; PMID:9784136	
A:Accession:	G71490	
A:Molecule type:	DNA	
A:Residues:	1-647 [ARN>	
A:Cross-references:	GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAC68226.1; PID:g332	
A:Experimental source:	serotype D, strain UM-3/Cx	
C:Comment:	This sequence was originally identified as homologous to part of a sequenc	
PIR:E72042).		
C:Genetics:		
A:Gene:	CT622	
Query Match	91.9%; Score 2943; DB 2; Length 647;	
Best Local Similarity	93.9%; Pred. No. 1.9e-117;	
Matches	613; Conservative 15; Mismatches 19; Indels 6; Gaps 1;	
QY	8 MMSGPEVSSNOSNMNPIINGQIASNSETKSTKASEASPSASSVSSVSLSSAKNALI 67	
DB	1 MMSGPEVSSNOSNMNPIINGQIASNSETKSTKASEASPSASSVSSVSLSSAKHALI 60	
QY	68 SLRDAIINKNSSPTDLSQLEASTSTSTVTRAAADYDAKSNFTAKSGLENATTLAY 127	
DB	61 SLRDAIINKNSSPTDLSQLEASTSTSTVTRAAADYDAKSNFTAKSGLENATTLAY 120	
QY	128 ETKMADIMAAIADMRLIANDSPSNHTEEVNIKAKLEAKOTJIKLNLVTLQONQSL 187	
DB	121 ETKMADIMAAIADMRLIANDSPSNHTEEVNIKAKLEAKOTJIKLNLVTLQONQSL 174	
QY	188 TEVLKTTSDADIPAINSOLEINKNSADQIINDLERONISYBAVLTNAGEVIKASSEAGI 247	
DB	175 KETLTTTSDADIPAINSOLEINKNSADQIINDLERONISYBAVLTNAGEVIKASSEAGI 234	
QY	248 KLGQALQSTVVDGDSQAQAVLQAOQNNNSPDNIAAKRELIDAAETVYNELKQHTGLTSP 307	
DB	235 KLGQALQSTVVDGDSQAQAVLQAOQNNNSPDNIAAKRELIDAAETVYNELKQHTGLTSP 294	
QY	308 LVKRAEEOISQAKDIOETIKPSGSDIPYGPSSAASAGSAGALSSNNSGRISLLDD 367	
DB	295 LVKRAEEOISQAKDIOETIKPSGSDIPYGPSSAASAGSAGALSSNNSGRISLLDD 354	
QY	368 VDNEAAIATLQCFRSMIDQFNVNPNATKAEIQAMEAQLTAMSDQLVGADGELPAETQAIK 427	
DB	355 VDNEAAIATLQCFRSMIDQFNVNPNATKAEIQAMEAQLTAMSDQLVGADGELPAETQAIK 414	

0y 428 DALAALKQPSNDGLATAMGVAPFAAAKVGSGSAGTACTVQMNKQLYKTFAPSSSSSY 487
 Db 415 DALAALKQPSNDGLATAMGVAPFAAAKVGSGSAGTACTVQMNKQLYKTFAPSSSSSY 474
 0y 488 AALSDGSAVKTLSLSESRGVSQSAISOTANPALRSRSRSGISEGGSADPSQAA 547
 Db 475 AALSDGSAVKTLSLSESRGVSQSAISOTANPALRSRSRSGISEGGSADPSQAA 534
 0y 548 ETIVADSGTLDGVYSRLQVLDLSMSTIVSNPQANOEIMOKLTATASISAKPOGYPAYONS 607
 Db 535 ETIVADSGTLDGVYSRLQVLDLSMSTIVSNPQANOEIMOKLTATASISAKPOGYPAYONS 594
 0y 608 ADSLQKFAAQLEREPVDSERSLAESQENAFKQAPFIOQVLYNATSLFSGTSL 660
 Db 595 ADSLQKFAAQLEREPVDSERSLAESQENAFKQAPFIOQVLYNATSLFSGTSL 647
 RESULT 2
 D72042
 conserved hypotheetical protein CP0018 [imported] - Chlamydia pneumoniae (strains CWI
 N:Alternate names: chlpn 76 kda homolog_1 (ct6523); hypotheetical protein CP0728
 C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: D72042; D81623
 R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, J.;
 Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606; PMID:10192388
 A:Accession: D72042
 A:Molecule type: DNA
 A:Residues: 1-651 <ARN>
 A:Cross-references: GB:AE001654; GB:AE001363; NID:94377031; PIDN:AD18867.1; PID:94377031
 A:Experimental source: Strain CWL029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: AB1500; MUID:20150255; PMID:10684935
 A:Accession: DB1623
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-651 <REA>
 A:Cross-references: GB:AE002165; GB:AE002161; NID:97188948; PIDN:AAF37914.1; PID:97188948
 A:Experimental source: strain AR39, HT cells
 C:Comment: This sequence was originally identified as homologous to part of a sequence (PIR:H71490).
 C:Genetics:
 A:Gene: CPn0728; CP0018
 Query Match 27.9%; Score 894; DB 2; Length 651;
 Best local similarity 34.6%; Pred. No. 8.3e-31;
 Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;
 0y 11 GPESVSSNOSSMNPITIOIASNSETEKSTRA-----SEASP--SASSSVSSWFLSSAK 63
 Db 6 GPSPIDETERRPPADLSAQGLEASANKSAEAQRITAGAEAKPKSEKTDTSVEMWSILRSAY 65
 0y 64 NALISLRD--ALINKNSPTDLSOLEASTSTSTYTRVAAKDYEAKSNFTANSGLENA 121
 Db 66 NALMSIAKLKGLASSNSSTSR--ADVDSFTATAPTEPPFPFDYKQAOATAYDTITS 124
 0y 122 KTLAYETKMDALMALDMMERLANSNHTTEVNNIKKALEOKPTID-----K 173
 Db 125 TSLADIQAAALVSLQDAVINIKDTATD-----ET-----AIAAEWETKNADAVKGAQ 173
 0y 174 LNKIVTLQNKSLTEVLTKTDSADQIPAINSOLEIRKNSADQIITKDERONI---SYEA 230
 Db 174 ITELAKYASNDQAALIDSLIGKLTSPFLQALALQSVANNKNAEILKEQMDNPVYVGGTTPA 233
 0y 231 VLTNAGEVITKASSEGIKLGALQSVIVADGDSQAAYVLOAQONNSPDNIATKELIDAE 290
 Db 234 IAAQSLVDDTQATATQIEKDGNAIRDAYFAGOVASCAVENAKSNNSITSIDSKAALIAAK 293

```

QY 291 TVNVLKQEHGTLTSPLYKKAEEQISQAQNDIQELKRS--GSDLTIVPSS--GSAASASSA 348
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 294 TQIAEAOOK---FPPSPITQIEAEQAVIQAEKPLKNIKPADGSDVNPPTGVGSGKQGGSS 350
QY 349 AGALSSNNNSGATISLLDDVDNEMAAIALQGRSRNIEQFNVNPNPTAELEAMEKQUTLAM 408
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 TGI-----RVSLDDAEHETASILMSGFRQMIHFNENPNPSQAQDELLAAQARRA 403
QY 409 SDQLVGADGELPAEIQAIKDALQAL--KQPSADGLATAMGOVAFAAKRVGGSGACTAGT 467
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 404 K---AAGDSSAAALADAQKALEALGKAGQGGQGITLNLGQIASAAVVSAGVPAASASI 460
QY 468 QMNVQOLYKTAFSSSSSYAALSDGYSAYKTLNLSYSESS--GVQASISQTPANPALSR 526
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 GSSVQOLYKT--SKSTGSDPKYQISAGYDAVYKSIDAYGARANDTRVINVSTPALTR 518
QY 527 SVSRGIESQGRSADASQRAETIVDSQTLDDVYSPRLQVLDLSLSTVTSNPOAQOEEIM 586
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 519 SVPRARTEKRG--PEKTDQALAVISGNSKRTLDGYISQVSALDSVQIITQSNPOANNEIR 577
QY 587 OKLTASISKAPQGPAYAVONSADSLSQKPAALEREFEVDGERSLSQENAFKQPAFTQQ 646
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 578 OKLTSAVTKPPQPGFYVQQLSNDSTQKFLAKLESFAGSRTAFRAEIKALSFETNLFIQ 637
QY 647 VLVNIASLFSGYL 659
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 638 VLVNIGSLYSGYL 650

RESULT 3
E86581
CHLPN 76 kDa homolog_1 (CT622) [imported] - Chlamydochila pneumoniae (strain J138)
C:Species: Chlamydochila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 sequence revision 02-Mar-2001 text_change 02-Mar-2001
C:Accession: E86581
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.;
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349; PMID:10811362
A:Accession: E86581
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-651 <STO>
A:Cross-references: GB:BA000008; NID:g8979100; PIDD:BA98935.1; GSPDB:GM00142
C:Genetic source: strain J138
A:Gene: CPJ0728

Query Match 27.9%; Score 894; DB 2; Length 651;
Best Local Similarity 34.6%; Pred. No. 8,3e-31;
Matches 233; Conservative 126; Mismatches 262; Indels 52; Gaps 17;

QY 11 GPESVSNMSSNNPIINGQIASNSETKESTKA-----SEASP--SASSSVSSWSSLASAK 63
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 6 GGPIDETERTPPADLSAQGLEASAAKNSAEQRIAGAEAKPKESKTDVSVERWSSILRSAY 65
QY 64 NALLISRD--ALLNNSSPTBLSOLEASTSTSYTRAAKYIDDAKSINFQAKSGLENA 121
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 66 NALMSIADKLGLASSNSSSSTRS--ADVDSITATATPTPPPTFDDYKIQQAQATAYDTFTS 124
QY 122 KTLAEYETFMADLMALQOMERLANSDSNNHTEEVNNNIKALEAQKQPTID-----K 173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 TSLADIQALVLSQADAVINIKPTAATD-----EET-----AIAAEWEYKKNDAVKVGAQ 173
QY 174 LNKLVTLQONKSLTEVLKTTDSDADQIPAINSOLEINKNSADQITKLERONI---STEA 230
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 ITELKAYASDNOAQALIDSGLKTLTSPDLLQALLQSVANNKKAELELKEKQMDNPVVGKTPA 233
QY 231 VLTNAGEVTKASSEGKIKIGQALQSLIVDAQDGSQAANVLAQOQNNSPDNINAAKELIDAE 290
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 234 IAOISLVDDQADATATQIEKDGNAIRDAYFRAGONASGAIVENAKSNNSISNIDSKAAALATPAK 293

```


Db 1604 -STSASGSASTSTSD---SRSTASASTSTMTSTSDOSMSLSTSTSTMSDSTSLSD 1659
QY 560 VVS-----RLOVLDs-----LMSTVSNPQANOEIMQKTLASI 593
Db 1660 SVSDSTSDSTASSTSGMSVSISSLDSTSTSTASSEVMSAISDSOSMSSEVNDSESVSE 1719
QY 594 SKA-----POFGYPAYONS-----ADSLQKFAQLEREFNGERSLAES 632
Db 1720 SNSEDSKMSGSGTSTVSDSGSLSVSTSLRKSESSESSSLSGSQSMDS 1768

RESULT 6
B89921
hypothetical protein ebha [imported] - staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: B89921
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: B89921
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-6713 <KUR>
A:Cross-references: GB:BA000018; PID:g13701232; PIDN:BA842527.1; GSPDB:GN00149
A:Experimental source: strain N315
A:Genetics: ebha

Query Match 8.1%; Score 261; DB 2; Length 6713;
Best Local Similarity 20.5%; Pred. No. 0.012;
Matches 150; Conservative 133; Mismatches 237; Indels 210; Gaps 31;

QY 10 GPESSVSSNOSSMNPITINGIASNETKESKASESPASSSVSSFLSSAKNALISL 69
Db 5610 NGDQNLANAKDKANAFVNSLNGLNQOQO-----DLAKHAIINNADTVSDVTDIYVNNQIDL 5663
QY 70 RDALINKNSPTDSLSL---EASTSTSTYTRAAKYDYDAKSNEFDPAKSGLEAKTILAE 126
Db 5664 NDAM-----ETLKLHVNEIPNAQOTVYONAD--DNAKTNEFDK-----5702
QY 127 YETKMADLALQDMERLANSNNTFEVNNIKKALEOKDTIDKLN-----KLV 178
Db 5703 -----RLANTLANSMT--NVNDINGAIQAVNDAIHLNNGDQRLQDAKDK 5745
QY 179 TLQONKSLTEVLKTTDSA-----DQIPAINSOLE-----INKNSADQITKDLERQ- 224
Db 5746 AIGSINGALANKLKEIRASNAITDQDKLIAKKAELANSIINNINKATSNQAVSOVOTAG 5805
QY 225 NISFEAVLTNAGEYIKASSSEAGIKLGOALOSIYDAGQSOAAVLQAOQNNSPNIAATKE 284
Db 5806 NHAIEQV--HANEPKAKIDANKDVQVQALIDEIDR-----NPNLTDK 5848
QY 285 LIDAEKRVNE-LKOEHTGLTDS---PLYKKAERQISOAKQDIOEIRPSGSDIPIVPSG 340
Db 5849 EKQALKRINQIIOGGHNDINNALTKEIEQAKQALQALQIDQIKDYKAKED-----5900
QY 341 SAASAGSAGALKSSNNSGRISLLDDVDN--EMAIALQGFMSIEQF-----NVNNP 392
Db 5901 -----AKQVDQVQVQALIDEIDONPNLTDEKQALKRINQIIOGGHNGINNA 5948
QY 393 ATAKELQAMEAQLT-----AMSQVQVQADGELPEIAQIAK 427
Db 5949 MTKKEIEQAKQALQALQALKEIKDYKAKENAKQDVQVQVQALIDEIDONPNLTDEKQALK 6008
QY 428 DALAQALQKQPSAGSLATAMQVAFAAKVGGSAGTAGTYOMNKKYKTAFTSSSSSY 487
Db 6009 DRINQIIOGGHND--INNAMTKKEIEQAK-----AQALQALQ-DIKDYK-----6050

QY 488 AALSDQYSAYKTLNLSYSESGVQSAISQTPANPALSRVSVSGIESQGRSADASQRAA 547
Db 6051 --AKEDAKNAIK--ALANAKRDQINS-----NPDLPEQAKAKAL-----KEIDEAEKRA 6095
QY 548 ETVRPSQTLGDVYSRQV--LDSLMSTIV-----SNQANO-----EETM-----586
Db 6096 LQVNEAQITDQNLNGLNGLDLIRNHWVEVDQPAVNEIFPATPQQLIVNCELLVHRD 6155
QY 587 -----OKTFASISKAPQFG-----YPAVQNSADSLQKFAQLEREFNGERSLAESQENA 636
Db 6156 DITTEDQILAHNLIDQLSLAEVIDTPTATISLSL---TAKVEVTLLDSKSVYVNPVKV 6212

QY 637 FRQPAFTQQ 646
Db 6213 VEKELSVYKQ 6222

RESULT 7
T12796

probable transglycosylase - Bacillus subtilis phage SPBC2
C:Species: Bacillus subtilis phage SPBC2
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999

C:Accession: T12796; A69911
R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauer, C.; Karamata, D.
submitted to the EMBL Data Library, August 1997
A:Description: The complete nucleotide sequence of the Bacillus subtilis spetac2 pro

A:Reference number: 217583
A:Accession: T12796
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2285 <LA2>

A:Cross-references: EMBL:AF020713; NID:93025478; PID:g3025510; PIDN:AAIC13005.1
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber
C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmertson, P.T.; Estlin, K.D.; Errington, J.; Febret, C.; Ferrari,
Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Gal
lechi, J.; Harwood, C.R.; Henalt, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Y, M.; Ogawa, K.; Ogilwara, A.; Oudeg, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowsky, A.; Se
akeuchi, M.; Yamakoshi, A.; Yanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69911
A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
A:Residues: 1-2285 <KUN>
A:Cross-references: GB:Z99115; GB:AL009126; NID:g2634478; PIDN:CAB14053.1; PID:e11835
A:Experimental source: strain 168
A:Genetics: yomi

Query Match 7.3%; Score 235; DB 2; Length 2285;
Best Local Similarity 19.3%; Pred. No. 0.034;
Matches 159; Conservative 141; Mismatches 312; Indels 210; Gaps 33;

QY 10 GPESSVSSNOSSMNPIT--NGQIASNSE-----TKESTASASPSASSSVSSWFLSS 61
Db 56 SAIDTYQKNLSYNOTVETSTVKNADGSVEKLTQYKRNKEITLORETKIINN-----109
QY 62 AKNALISLRALINKNSPDSLSQLEASSTSTVTVAAKVDKDEAKSND-----TAK 115
Db 110 -RNTALQETQEVNKLQATQETKLGQVQKTYQNRMLQGPTRVYQKNHNGDDIYVTTPDK 168
QY 116 SGLENAKTLEVEYKMDLMAALQDMERL-----ANSPSNNHTEVYN 158
Db 169 TNSTSKTTNTYDQQRRAIQKQLEKLRQOGIYVDTTITSSIGRKINTAQSAQIEALQ 228


```

Db 305 QLTSTRKDSKNSVTDSSNKSDIESANTELKQALAKANADKVDNLAKSKEIQLNNSVNA 364
Qy 122 KTL-AEYETKMAIDLMAALODMER---LANSDPNNHTEEVNNIKKALEAKODTIDKLU 177
Db 365 NTLISALTLTKDNTIQAKTELEKEVQKADQAIKSNNTASQSKSLDQKVAITK--KL 422
Qy 178 VTLQONKSLTEVLTQTTSDADQIPAINSOLEINKNSAD--QIKLERONISTEAVL--T 233
Db 423 ETEFNKKEKKEFNELKOTRN-----QIOEFINTKNNPNYSELISQLTSTRKDSKNSVTDSS 477
Qy 234 NGEVTKASSEAGIKIGQALQSTIVDAGDQQAVALQAOONSPDN--AATKELIDAEFK 292
Db 478 NMSDISANTELKQALAKANADKVDNLAKS--IKELQNNSVNANTLSAKLTD-----K 531
Qy 293 VNELKOEHTGLTDSPLVKAEEOISOAKDIOEIKPSGSDPIVPGSGSASAGSAGAL 352
Db 532 DNTIQAKTEL-----EKEVQKADQAIK-----SNNTASQSKSSL 568
Qy 353 KSSNNSGRISLLIDVDNEMAAIALOGFRSMIEQF---NVNNPATAKELQAMEAQLTAMS 409
Db 569 DAKVAETTKLETFNFKDKEKKEFNELKOTRNQIOEFINTKNNPN-----NYSSELISQLTSTR 624
Qy 410 DQVLG-ADGELPAEIOAIDALQALQKPSADGLATMGQVAFNAKVGSGSAGTAGYQ 468
Db 625 DSKNSVTDSSNKSDIESANTELKQALAKANADKVDQ--DNLAKSKEIQLNNSVNA 662
Qy 469 MNV-----KQLYK-----TAFSTSSSSYAALSDQYSAYKTLN-- 502
Db 683 AKLTQDNTIQAKTELEKEVQKADQAIKSNNTASQSKSLDQKVAITKLETFNPKD 742
Qy 503 -----SLYESRSGVOSAISQTA--NPALSRVSRSIGIESQGRSADASQRAETIVDSOT 556
Db 743 KEAKFELKOTRNQIOEFINTKNNPNYSELISQLT-----TSKRDSKNSVTDSSM 792
Qy 557 LQGVSRLOVLSIMSTIYSNPQANOEIMQKLTASISKAPORGYAVONASADSLQKFA 616
Db 793 KSDIESANTELKQALAKANADKVDNLAKS--IKELQNNSVNANTLSAKLTD-----K 848
Qy 617 QLEREFVDEGRSLAE--SOENAFRQOPAFIOQVYL 648
Db 849 KLETT-----KNKLAEEITKADAIKNNPSSKQAL 878

```

RESULT 10

```

F84730
probable myosin heavy chain [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84730
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: F84730
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1269 <SNO>
A:Cross-references: GB:AE002093; NID:g6598483; PIDN:AAC69932.2; GSPDB:GN00139
C:Genetics:
A:Gene: Atg932240
A:Map position: 2

```

```

Query Match 7.1%; Score 226.5; DB 2; Length 1269;
Best Local Similarity 20.8%; Pred. No. 0.035;
Matches 145; Conservative 142; Mismatches 286; Indels 125; Gaps 28;
Qy 8 MEGSPSVSSNOSNMPI-----INGOIASNETKESTKSPASPSASSSVSSWSPLSS 61
Db 242 LKTKSKAKEMEKMASTLOEIKELNKKSENKVEALAKSSAGELAAVQEE-----LAL 296
Qy 62 AKNAL-----ISLRDAIILNKNSSPTDSLQLEASTSTSTVTREAAKADYDEAKSNFDYAK 115

```

```

Db 297 SKSRLETEQKVSSTELALIDE---LTQLEQKAKASERKEELSVLQDDAQTKQALU 353
Qy 116 SGLENAKTLAEYETKMAIDLMAAL-ODMERLANSDPNNHTEEVNNIKKALEAKODTIDKU 174
Db 354 SEQEGINSKLAELKEKLELESKQDEKLR--ANEKLAELKKEKLEA--NVAEYT 409
Qy 175 NKLVTLOONKSLTEVLTQTTD--SADQIPAINSOLEINKNSADQIINKLERONISTEAVLT 233
Db 410 SNVAIVTEVCNLEELKLTSDENFSKTDALLSQALSNNSLEQKLSBELH-----S 462
Qy 234 NGEVTKASSEAGIKIGQALQSTIVDAGDQQAVALQAOONNS-----PDNIAT 282
Db 463 EAGSAAATAATQKNLELEDVVRSSQAEEAKSQIKLETKFTAAEQKAMELEQQLNLQ 522
Qy 283 KELIDAEFKVNLKOEHTGLTDSPLVKAEEOISOAKDIOEIKPSGSDPIVPGSGSA 342
Db 523 KS--SDAEKELKESKSELOTA--IEVAEEKKATITQMDIEYKQKASEL----- 569
Qy 343 ASAGSAGALKSSNNSGRISLLIDVDNEMAAIALOGFRSMIEQFNVNNPATAKELQAME 402
Db 570 -----ELSLTQSSARNSELEDL-----RIALQKGAHEHEDRANTHTQ-----RSIE 610
Qy 403 AQLTMSDQLVGADG-----LPAP---IQAIKDLAOLK---QPSADGLATANG 447
Db 611 LBGICSSQSKHEDAGGRUKDELTLQTEKYRIQLEDEQVSSLEKKGTEADSKG-YLG 669
Qy 448 QVA-----FAAAKVGSGSAGTAGYOM--NVKOLYKTAFTSTSSSSYAALSDQYSAYKTL 501
Db 670 QVAELQSTLEAFQVKKSSLEALNLTENKELTEMLNMTSKKKLEATVDIYSV----- 725
Qy 502 NSLYSESRGVOSA-----ISQTPALSRVSRSIGIESQGRSADASQRAETIVDSOT 556
Db 726 --KISESENLEESIRNELNLTQKLESIENDLKAQIQ--ESEVEXLKSAESELQKGRE 782
Qy 557 LQGVSRLOVLSIMSTIYSNPQANOEIMQKLTASISKAPORGYAVONASADSLQKFA 616
Db 783 IDEATTKRMELEALHQSLSDISEHRLQKAMEEPTSDSE-----ASSLTEKLR 830
Qy 617 QLEREFVDEGRSLAE--SOENAFRQOPAFIOQVYNVNAS 653
Db 831 DLGKIKSYEQLAEASGKSSSLKEK--LEQTLGRLLA 866

```

RESULT 11

```

JG6009
surface-located membrane protein Imp3 precursor - Mycoplasma hominis
C:Species: Mycoplasma hominis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: JG6009
R:Ladefoged, S.A.; Jensen, L.T.; Brock, B.; Birkelund, S.; Christiansen, G.
J. Bacteriol. 178, 2775-2784, 1996
A>Title: Analysis of 0.5-kilobase-pair repeats in the Mycoplasma hominis Imp gene sys
A:Reference number: JG6009; MUID:96213016; PMID:8631664
A:Accession: JG6009
A:Molecule type: DNA
A:Residues: 1-1302 <LAD>
A:Cross-references: EMBL:X95601; NID:g1197335; PIDN:CAA64858.1; PID:g1197336
C:Genetics:
A:Gene: Imp3
A:Genetic code: SGC3
C:Superfamily: surface-located membrane protein Imp3; tetratricopeptide repeat homolo
C:Keywords: duplication; membrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-1302/Product: surface-located membrane protein Imp3 #status predicted <MAT>
F:957-992/Domain: tetratricopeptide repeat homology <TT1>
F:993-1026/Domain: tetratricopeptide repeat homology <TT2>
F:1089-1120/Domain: tetratricopeptide repeat homology <TT3>
F:1154-1190/Domain: tetratricopeptide repeat homology <TT4>

```

```

Query Match 7.0%; Score 225.5; DB 1; Length 1302;
Best Local Similarity 21.6%; Pred. No. 0.04;
Matches 150; Conservative 139; Mismatches 223; Indels 181; Gaps 38;

```


Db 1947 DKTASLNTQTHDDVHPIKPKDAEKTINDDLARVATLQVNRKYSNNRKADALKAITATL 2006
QY 466 TVQNMVVKQLYTAFSTSTSSS-----YAAALSDGYSAYKTYLNSLYSESSGVSQSAISOAN 521
Db 2007 KLQMD--BELKAFARNADVADLAKRFNVALSD-----IEAVITTEKEN 2046
QY 522 PALSRVSRSRGIESGRSADASQRAAETIVRDSOTLGDVYSRLQVLSLMTSTYSENPAN 581
Db 2047 SLKRDINAQOTYAKFKAIATPEQJLAKYKVLIDQYVAD-GNNMIDEDATINDIKQHTQFI 2105
QY 562 QOEIMQ-KLTASISK-----APQFGYPAYQNSADSLQKFAQLEREFDVG 625
Db 2106 VDEILAIKLPAAETKVPSPKEIQPAPKVCPTPIKKEETHESRKVEKELPMTGSEG 2158

RESULT 19
T13030
microtubule binding protein D-ClIP-190 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13030
R:Lantz, V.A.; Miller, K.G.
J. Cell Biol. 140, 897-910, 1998
A:Title: A class VI unconventional myosin is associated with a homologue of a microtubul
A:Reference number: 217588; MUID:98139549; PMID:9472041
A:Accession: T13030
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1690 <LAN>
A:Cross-references: EMBL:AF041382; NID:g2773362; PID:g2773363; PIDN:AA896783.1
A:Experimental source: strain Oregon R
C:Genetics:
A:Cross-references: FlyBase:FBgn0020503
C:Keywords: cytoskeleton

Query Match 6.6%; Score 212; DB 2; Length 1690;
Best local similarity 19.3%; Pred. No. 0.21;
Matches 136; Conservative 133; Mismatches 300; Indels 134; Gaps 26;

QY 4 HHHHESCPESYSSNQSSMNPITNGQIASNSE-----TKESTYASESPSSASSVSS- 55
Db 929 YHTDVESTKQLEAANMALEKYNKEYSRAESDADLQDKVEITDTLHAELQAEKSSSSA 988
QY 56 -----WSFLSSAKNMLISLRDA-----ILNKNSSPDTLSQLASHTSTVTVAA-- 101
Db 969 LHTKLSKSDETATGHKELTSKADAMSOEMLQKELELQELRQLODSDOSQTKLAEGER 1048
QY 102 --KDYDEAKSNF--DTAKSGLENAKTLAEYETKMDLM-----AALQDMERLANSPS 150
Db 1049 KEKSFEEIKNLQEEVTKAKTENLELSTGTQTTIKDLOERLEITNAELQHKEMASSEDQ 1108
QY 151 -----NNHTEEVNNIKKALEAKQDPTIDKLNKLVTLQONKNSLTEVL 191
Db 1109 KIADKLTVEAIOVANANISATINAEIATVLEVLQAEKSETNHIFELFEMEADNMSERLIE 1168
QY 192 KTTSDADQIPAINSOLEINKNSADQIHKDLERONISYEAVNLNAGEVIRKASEPAIKLQ 251
Db 1169 KYTGKEELKETHQLDLDRQKKFELEELKQAOQSEQKLOES---QTSKELTETIQ 1224
QY 252 ALOSLIVDAGDSQAAVLQAOQNSPDNIATKELIDAETKVENELKQEHGTGLTDSPLVK 311
Db 1225 SLQELQDSVKQKEELY-----QNLKRESSSITIEAONTKINESNOLEKNTS--CKE 1277
QY 312 AEOISQAKDIOETIKPSGSDPIVPGSAAASAGSAGALKSSNNSGRISLLDDVDNE 371
Db 1278 TODOLLESQKREKQLOEBAKL-----SGELQOVQEANGDKIDS--LVKVEELVAVLEEK 1330
QY 372 MAAT-----ALQGFMSIMEQFVNNPATAKELQAMEAQLTAMSDQVAGADGLPAIEQAI 426
Db 1331 LQATSQLDQAOATNKELQELLVKSOENEGNIGSGLAVTEKLODLEQANGEL----- 1383
QY 427 KDALQALKOPSDADGLATAMGQVAFPAAKVGGSGAGTAGTVQNMVKQLYKTAFFSSTSSS 486

Db 1384 KEALQ-----KENCKLELQGLK-----DESNVLYESQK-----SHNE 1417
QY 467 YAAALSDGYSAYKTYLNSLYSESSRGVSQSAISO--TANALSRVSRSRG-IESQGRSADAS 543
Db 1418 IQDKLEQAQOQKERTL-----QOETSRLAEQLSOLQANBELKSLQOOLLEKGEFPTQ 1473
QY 544 QRAAETIVRDSOTLGDVYSRLQVLSLMTSTI-----VSNPOANO-----EETIMOKL- 589
Db 1474 LAEYQKVIDEMDDAASVSAL--LEQLONRVVAELETALRQANDQAKTAVLETKELRQLE 1531
QY 590 TASISKAPQ-FCYPAYQNSADSLQKFAQLEREFDVSGRSIAE 631
Db 1532 SLELEKSRREVLSLKQNMGASS--RSKGDEVESLIDFTSLAK 1572

RESULT 20
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abse
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SM1>
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 6.6%; Score 211.5; DB 2; Length 1822;
Best local similarity 19.6%; Pred. No. 0.24;
Matches 147; Conservative 131; Mismatches 300; Indels 171; Gaps 30;

QY 15 VSSNQSNNPITNGQIASNSESTKAS-EASPASSSVSSWSFLSSAKNMLISLRDAI 73
Db 902 IAYNEDVINAQOLDLNLNLEKDSEETKAAIDANPMLTPEERA-----KAIAVEELY 953
QY 74 LNKN-----SSPTDSLQLEASTSTSTVTRV-----AAKDYDEAKSNF--DTAKSG 117
Db 954 NNAESDILSKPTPEYVQAVEKADKDLAKVELQAAADGAKGIEANPMLTPEEKDVAKKA 1013
QY 118 LEN-----AKTLAEYETKMDLMAALQDME-RLANSDPSNNHTEEVNNIKKAL- 164
Db 1014 VEDAVKAVTADIDKASTPEVDFATSDGVKALDAEEFKATQDANKKIAKEAESAKKAID 1073
QY 165 -----EAQKDTIDKLNKLVTLQ-----NQKNSLTEVLKTTSDADQIPA 202
Db 1074 DNPMLTPEEKESAKNAVEAKVATPAIDKASTPDAVOVEEDKGVAAINLITAKADAGV 1133
QY 203 INSQLEINKNSADQIHKDLERONISYEAV---LTNAGEVIRKASEAGIKLQALQSIVD 258
Db 1134 IAAKL-----ADEIKKLEDKQAEKAIIDASTMTNEEKATIAK-----KALQDVVD 1178
QY 259 AG-----DOSQAAYLQAOQNSPDNIAT-----KELIDAE----- 290
Db 1179 KGKAELEDAARVATNEIEHATTEKAKAEELAGEKSLTDTGTEADDAVELAKKELAKEA 1238
QY 291 -----TKVNELKQEHNT--GLTDP-----LVKKAEOHSO-----AQDI 323
Db 1239 IRTEEBEATKIVKLAETTKRAIENPNLSDCKAEIKKLTDAVAKLATIRDAKDKRT 1296
QY 324 QETKPSGSDPIVPGSAAASAGSAA-GALKSSNNSGRISLLDDVDNEMAIALQGRS 382
Db 1299 QEAERKAQLADLEKAKETQIADKKAIDRLFTLVKDGELKATQKAKKIAKDAADAAKEA 1358
QY 383 MTEQFVNNPATAKELQAMEAQLTAMSDQVAGADGLPAIEQATIDALQALKOPSAGCL 442
Db 1359 IASNPNLTDAEKKTFTTDDVDAEVAAKANDAIISAATS--PADVOKEEDDAVVAIAEDVDA 1416
QY 443 -ATAMGQVA--FAAKVGGSGAGTAGTVQNMVKQLYKTAFFSSTSSSAAALSDGYSAYK 499

Db 1417 KODAKNKIKADAAAKAIGS-----NPNLDAEKJFTTDAVDAEVAAK-NDATSAAT 1468
QY 500 TUNSLSESRSQVS-----AISOTANDALSR--SVSRSGIESQGRASDAQRAETI 550
Db 1469 SPADVQKEEDAGYAAIAEDYDLAAKODAKNKIKESDASAIADANPNLTDAKESAKKA 1528
QY 551 VRDSQTLGDVYSRLQVLDLSMTSTIVSNPOANOEIMOKLTASISKAPQEGYPAVONSADS 610
Db 1529 VD-----ADAKKATDAIDA--STSPVEAQSAEDKGVGSIADQVLDLA-----AKODAKNK 1575
QY 611 LQKFAQLEREPYDGERSTAESOEANFRK 639
Db 1576 IAKFAA-AAKEAIDANPNLSDAKESAKR 1603

RESULT 21

FmbB protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: D90011
R:Kurda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-Ui, T.; Kobayashi, N.; Savano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: D90011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2481 <KOR>
A:Cross-references: GB:BA000018; PID:g1j3701961; PIDN:BAW43253.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: fmbB(mrp)

Query Match 6.6%; Score 210.5; DB 2; Length 2481;
Best Local Similarity 19.6%; Pred. No. 0.41;
Matches 153; Conservative 132; Mismatches 313; Indels 184; Gaps 32;

QY 19 QSSMNLINQ-----IASNSEKSTKASEASPSASSSSVSFLSSAKNALISLDAI 73
Db 659 QDVITPVAKPOAKODITIAVTRKQIKSNASLQDEKDVANDKIGKIEKAIKIDAAI 718
QY 74 LNK--NSSPDSLQLEASTSTSTVTRVAAKDYDE-----AKSNFDPAK-----SGLE 119
Db 719 TNGQVEAIKTKAINDINOTTPATKAALAEEDDEVYQAIIDAPLNPOTITNEEVAATE 778
QY 120 NAKTLAEYETKMADLMAALQDMERLANSDBSNHTEEVNNIKALEAQKDITDKLKLVT 179
Db 779 RINAAYSGVKAIEATTTADLERKNEE-----IFKIENTDSTQTKMAYKEVRQAAT 833
QY 180 LQN-QKSLT-----EVLK-----TDSAQIPAINQLEINKNSAQIITKDERON 225
Db 834 ARAAQNAVYSNATDEEVAENAAVDAQAQTEGLDIOVYKSOQEV---ADTKAKVLDKIN 889
QY 226 -ISYEAVLTNAGEV-----IKASSEAGIKGLQALOSIVDAGDSQAVALQAQ 272
Db 890 AITQQAIVKRAATDEVENAYNTKQEIQNSNASTTEKEKAYITELDAKKOBEATINDAAN 949
QY 273 NNS-----PDNIAATKELIDAAETKYNELKOEHT-----GLTDSPLVKAEEQT 316
Db 950 TNSDVTAKKNGIAAINQVQAATTKKSDAKAEIAQKASEKTAIEAMNDS-----TTEEQ 1005
QY 317 SQAKQIOIEIKPBGSDIPIYVPGSAAAGSAGALKSSNSRISLILDDVDNEMAATA 376
Db 1006 AAKDKVQDAQVAVTANADI-----DNATANTDVNNAKTTEA--TIAITPDAVKKPA--A 1055
QY 377 LOGFRSMIE-----QFVNNNPATAKELQAMEAOLTAUSDOLVGADGELPA-----ETQAIK 427
Db 1056 KQAIADKVAQOETAIDANNNGSTTEKEAAKQO--VQTEKTAADAALIDAASHVNEVAEK 1112

QY 428 DALAQALK--QPS-----ADGLATAMQV-----AFPAKVGGSAGT-- 463
Db 1113 NAEIAKIEALQPATTTKDKNAKQAIATFKANERKTAIAQOTDITAEETIAAANADVNAVQA 1172
QY 464 -----AGTYQMANKQLYKPAFFSTSSSSYAAALSDQYSAKYKTINSLYSESRSQVQAIQ 518
Db 1173 NSNIEANSQNDVQAKTT--GETSIDQYTPYNNKATARNETAILNKNKLOEIQTTPDA 1230
QY 519 T-----ANPALSRSVSRSGIESQGRASDAQRAETIYRDSQTLGDYV 561
Db 1231 TDEKQADAEANTENGKANKQAIASATITNAQVDEAKANNEAATINATPPYVKKQAAKDEL 1290
QY 562 SRLQVLDLSMTSTVSNPQ-----ANDEIMOKLTASISKAP----- 597
Db 1291 DQLQ-----ATQTNVINNDQATNNEKEBAIQOALATAVTADAKNNITTAATDNGVDTAKDAG 1347
QY 598 ---QFGYF--ANQNSADSLQKFAQLEREVDGERSLAESQENAFR-----KQAPFIQ 646
Db 1348 NSIQSTQPATVAVKSNAKNEVDQAVTTQNOAIDNTTGATTEEKNAARDLVAKAEKAY -QD 1406
QY 647 VL 648
Db 1407 IL 1408

RESULT 22

T34107
hypothetical protein c18C4.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34107
R:Gatlung, S.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid C18C4.
A:Reference number: Z21478
A:Accession: T34107
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1091 <GAT>
A:Cross-references: EMBL:U55369; PIDN:AAC25825.1; GSPDB:GN00023; CESP:C18C4.5
A:Experimental source: strain Bristol N2; clone C18C4
C:Genetics:
A:Gene: CESP:C18C4.5
A:Map position: 5
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2;

Query Match 6.6%; Score 210; DB 2; Length 1091;
Best Local Similarity 19.0%; Pred. No. 0.14;
Matches 136; Conservative 141; Mismatches 289; Indels 150; Gaps 25;

QY 8 MESGPESVSSNOSMNPITNGQIASNSETEKSTKASEASPS-----ASSVSSMSFLS 60
Db 108 MEQEIENLQROLNITKTASIQSLMLAKSDSSKTKDLSEBENTLKLKVEDIQKQVSPMSQM 167
QY 61 SAKNALI-SLRDAIILKNSSPDSLQLEASTSTSTVTRVAAKDYDEAKSNEDTAKSGLE 119
Db 168 QDNKSEIQKMDAISYNDVSRQNMDSLSEKLSMDRTLREDOQKQSLQSQTETLKNALS 227
QY 120 NAKT-----LAEYETKADL-----MALQDMERLANSDBSN- 152
Db 228 TSESTSLMLKDKLAQEQNALDLKNEANQKTSRISLIFESGRIRKELQOALSDEKDNNA 287
QY 153 -----HTEEVNNIK-----KALEAQKDITDKLKLVLQ---NOKNSLFEVLKT 193
Db 288 ILNVQUREKDGKIDRIQVLDLAAESRAQAEEDYVRMKERITISKKDDSNLNLQDELRL 347
QY 194 TDSADQIPAINQLEINKNSAQIITKDERONTSYRAVLTNAGEVYKASSEAGIKIGQAL 253
Db 348 TEEKYQ-----QAQKIEMLDETIKQOETQ-----IRDLGRSLDEAK 384
QY 254 QSIYVDAGDSQAANVLAQOQNSPDNI--AATKELIDAAETKYNELKOEHTGLTDSPLVK 311
Db 385 RQLOKMSQORQNEVARQGEDSARSMEKATKEIKKLSQV-QLOQOLE--QDELEQK 441


```

Db 630 IELSEVKIDELQSSIEVLSTKLEREYSSNLQNEELKLSLRFNEELQADLMSKAKNEELE 689
Oy 123 TLAE-----YETKADLMAALQDMERLANSDESNHTEE--VN 158
Db 690 QOIESSSEFVSITASKEMKLMKMSSEKQMSSEIMASLAFQEMQSTQADVAASEDKVK 749
Oy 159 NIKKALEAKOKTIDKLKLVTLQONKSLTEVLKTTDSADQIPAINSOLEIKNS---AD 215
Db 750 QVESLLENKLEPEELNNLRA-----NLKDSMDKTLIDQSDLELQOQSSDLAD 797
Oy 216 QIKDLERONISYEAVLTNAGE-----VIKASSEAGIKLQALQSLIVDAGD 261
Db 798 RLQEDLRTSDARVQELINQVSELBOQIEVSSREFVSITFANKEMQKLQSS--EAQISEMT 856
Oy 262 QSOAVALQOONSPDNIAA---TKELIDAAET-----KYNELK- 297
Db 857 ASLTFQEMQSTRADVAASEDKVKELESLEENLEKEPELELNPLANLKSNGKYLELQS 916
Oy 298 -----QEHRTGLTD-----SPLYKKAEBQISQAOKDIQIKPSGSDIPVPGSGSA 343
Db 917 QLDLAQOQSDLTDLQOEDLRTSDARVQELINQVSDQSELETAQODTNAGV--MEALKS 975
Oy 344 SAGSAGALKSS-----NNSGRISLDDVDNEMAALQGRSMIEQFNANNPATKEL 398
Db 976 EGGESEYALRAELDAVQEKGRSSDLVTSLECK-----IQELETAIESSTAENQKSKTI 1030
Oy 399 Q-----AMEA-----QLTAMSDQLVGADGELPAEIOAIKDAL 430
Db 1031 QDFTKVSLESOICELKSQNBQMEIDTNLMDOLEMSQLESANAEI----- 1079
Oy 431 AQALKPSADGLATAMGOV-----AFAAKVGGSGAGTGTV-----QKN 470
Db 1080 -IELTRTSAETIDKLREGYEKSTKAMQOEHLAEIVAKIESRDVENADQAKHKEEDER 1138
Oy 471 VKQLYKFAFSSTSSSYAALSDGYSAYKTLNLSYSEKSGV--QSAISQTPANPALSR- 526
Db 1139 LQSVIDTLRTSQSTIEESQAKSE-----ELNSRIKELQASIEFQKALADTENKQEKV 1192
Oy 527 -----SYSRGIESQGRSADASQRAAETIVRDSQTLGDIVSRLOVLD 569
Db 1193 ELEKVOBQMLNLVQAFEFKESIRLEWNS--SLSNANKELEAAEELASQKENTIVTLES 1250
Oy 570 LMSTIVSPQANQOEIMQKLTASISKAPQFGYPAYONSADSLQKFAAOLREFVNGERSL 629
Db 1251 RIETISQOPEARLEE-----ANVKKSOAMQGLTQSLQSOROLGEMHEKMEASDRKV 1304
Oy 630 AESQENA 636
Db 1305 IEVEBOA 1311

```

RESULT 27

C89921
 hypotheetical protein ebh [Imported] - Staphylococcus aureus (strain N315)
 C:Species: Staphylococcus aureus
 C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C:Accession: C89921
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogura, A.; Mizutani, O.; Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirata, K.
 Lancet 357, 1225-1240, 2001
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A:Reference number: A89758; MUID:21311952; PMID:11418146
 A:Accession: C89921
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-3890 <KUR>
 A:Cross-references: GB:BA000018; PID:g13701233; PIDN:BA042528.1; GSPDB:GN00149
 A:Experimental source: strain N315
 C:Genetics:
 A:Gene: ebh

Query Match 6.5%; Score 208; DB 2; Length 3890;
 Best Local Similarity 20.0%; Pred. No. 0.96;
 Matches 152; Conservative 123; Mismatches 286; Indels 200; Gaps 32;

```

Oy 6 HHMEGPPSVSSSSNMPIINGOIASNSETKSEKASEAFSSASSVSSMSFLSSAKNA 65
Db 2953 HSIDTQSTTGCTQOSINAYNAKLTAAKRVQOINQVLAGSPPTQDQINTNTSAANOAKSD 3012
Oy 66 LISLDAILNKSSPTPDSLSQLEAS-----TSTSTYTRAAKY----- 104
Db 3013 LDHARQALTPDKAPILQNAQTQLEOSINOPTDTGCTTASLNAVYNNKLOAAROKLTEINOV 3072
Oy 105 -----DEAKSNFDTAKSGLENATLAEYETKADLMAALQDMERLANS 147
Db 3073 LINGPTVONINDKVAENQAKQOLTARQGL-----TL-----DROPLATTLTGASNL 3120
Oy 148 DPS--NNHTEEVNNIRK--ALEAKDTIDKLKLVTLQONKSLTEVLKTTDS--ADOIPA 202
Db 3121 NQAQONNFTQOINAAQNHAALETIKSNITALTMTATKLDVADNNTIKSGQNTYDAPPA 3180
Oy 203 INSQLEIKNSADQIITKLERONISYEAVLTNAGEVITASSFA-----GIKLG--QALQS 255
Db 3181 NKQAYDNNVNAKGVIGETTFPTMDVNTVNOKAAVY--KSTKDALDGOONLQRAKTEATPA 3239
Oy 256 IVDAGDQSOAAV--LQAOONSPDNIAAT---KELIDAAETKYNELKO--EHTGLTDSPL 308
Db 3240 ITHASDLNQAKNALTOOVNSAQVAVNDIKOTTQSLTATATGTLKRGVANHNOVQSDN 3299
Oy 309 VKKAE-----EQISOAKDIOETIKPSGSDIPVPGSGSASAGSAGSANNNSGRISL 364
Db 3300 YVNADTNKKNDVNNAYNNHANDIINGNAQHPVITPDSVNNALSNVTSKEHALNGEAKLVNA 3359
Oy 365 LDDVNEMAAIALQGRFSMIEQFNANNPATKELQAMEQOLTAMSDQLVGADGELPAEIQ 424.
Db 3360 KQEAFTALGHNL-----NLNN-----VQRONLOSQINGAR---QID 3392
Oy 425 AIKDALQALQKPSADGLATAMGOVAFAPAAKVGSGSAGTAVQNNVQOLYRTAFSSSTSS 484
Db 3393 AV-----NTIKQ--NATNLNLSAMGNLRQAV-----DKQVYKRT----- 3424
Oy 485 SSVAAALSDGYSAYTTLNLSYSEKSGVQSAISQTPANPALSR-----RSVSR----- 531
Db 3425 EDYADADTAKQNAV---NSAVSSA---ETIINGFANPTMSVDVNRATSAVTTKKNALN 3477
Oy 532 GIESQGRSADASQRAAETIVR--DSQTLGDIVSRLOVLDL-----MSITVSNP 578
Db 3478 GDEKLVQSKTDAKRAIDALPHLNNQAKADVSKINAASNIAGVNTVKQOGTDLNTRAMGNL 3537
Oy 579 QA--NOEIMQKLTASISKAPQFGYP-----AVONSADSLQKFAOLE----- 619
Db 3538 QCAINDEQ-----TTLNSQONTQDATPFSKRTATYNVQAAKQDLNKSNGONKTKQDVTEAM 3592
Oy 620 -----REFVDGERSLAEQENAFKRPAPFIQOVLVNIASL 654
Db 3593 NOVNSAKNNLDQTRILLDQAKQTA-----KQOLNNMTHL 3625

```

RESULT 28

C82206
 methyl-accepting chemotaxis protein VC1403 [Imported] - Vibrio cholerae (strain N1696)
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
 C:Accession: C82206
 R:Heidelberg, J.F.; Eissen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, L.; R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: C82206
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-641 <HEI>

A:Cross-references: GB:AE004218; GB:AE003852; NID:96555881; PIDN:AAF94560.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype EI Tor
C:Genetics:
A:Gene: VC1403
A:Map position: 1

Query Match 6.5%; Score 207.5; DB 2; Length 641;
Best Local Similarity 21.6%; Pred. No. 0.088;
Matches 148; Conservative 105; Mismatches 252; Indels 179; Gaps 31;

16 SSNOSSMNPINQIANSFETKESKASESPASSSVSSWPLSSAKNALISLRPAILN 75
DB 9 TSSSSFSF-----SNHSTSNAPKKEAD-----LRQAKTLAKKQQAER 50
QY 76 KNSPTDSLSQLEASTSTSTVTR-----VAAKDYDEAKSNPDYAKSGLEN-AKTAEY 127
DB 51 VASATAEELLAGIESSSAMTQLRSSMEQIAYGAESSSANTKQSEYAVSQNNNTINQATY 110
QY 128 ETKMADLMAAL-QDMERLANSPPSNHTEEVNKKALEAKQKPTIDKLNKLVTLQNO--- 183
DB 111 ATOSKOILTKLEQDI-----VDVANKINEMVSNVQTSERQNDV---KRMIELSEQAAK 162
QY 184 -NKSLEVLTKTDS-----ADQIPAINSOLEIKKNSADQI 217
DB 163 INDVAVQVIHIDQTLNLNAAIEGRGKGGKGFVAVADTVRTIAEKAKNAANIESL 222
QY 218 IKDLERONISYEAVLTNAGEVIRKAS---EAGIKLGOAL-----QSTIVD----- 258
DB 223 IKD-----ISNGAEVMSNG--VKSSQKAKSEYQKHVYVKKOITLIRNEVVDLYKXSEVL 275
QY 259 ---AGDQSOQAVIQAQNNSPNIAATKELIDAEKRVNELKQEHGTL----- 303
DB 276 LKADEMSSAAVATALGSESVSQ--AAEQSACESESLSDQOQIALDGAVTAAGSLDE 333
QY 304 -----TDSPLVKKAREQISOAQK---DIOETKPSGSDPIPVSPSSASAGSAGAAKLS 354
DB 334 LIDELTSTDIYKSEVAAAEELASGLEIRSSNET---KALNOISSGAGHMAK 388
QY 355 SNNSGRISLLDD---VDNEMAAIALOGFRSMIEQFNVPNPATKELQAMEQLTAMSD 410
DB 389 SVETGITSLSQIEQARGLKERSESSLKACEKNLTGIEBKKTVDMEILAITSTTAATE 448
QY 411 QLVGADGELPAETQAIKDALQALKOPSDAGLTATAGQV--APFAAKVGGSGAGTACTYO 468
DB 449 NL-----NEM-ANIERISROIIDKIVDGISVSIQTAMLAVGAEEARAGEYGGKFA-VVS 502
QY 469 MNYKOLYKTA-----FSSSTSSSYAALSNGYCAKKT 500
DB 503 TDIQNLANDAEVAEQIKQVKNIOEQINIVRKDLADILSTVMEEQKALVT-----IKO 557
QY 501 LNSLYSESRGV--QSAISOTANPALRSYV--SRSGIESQGRSADASQRA---AETIVR- 552
DB 558 LDVVRMSQSVYLGSKRISESAG-GIERSLADARAGMQQATATAEESSHATGTGAATRAARQ 616
QY 553 DSQTLGDVYSRLD---VLDLSMS 572
DB 617 QSSSTSELASAIENIAVADELQS 640

RESULT 29
B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: B43402; A43402
R:Takahashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
A:Reference number: A43402; MUID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA

A:Residues: 1-2007 <TA>

A:Cross-references: GB:M93676; NID:9212448; PIDN:AAA48988.1; PID:9212452

A>Note: the sequence of residues 212-221 and 632-652 and the corresponding nucleotide

A:Accession: A43402

A:Molecule type: mRNA

A:Residues: 1-211;222-631;653-2007 <TA2>

A:Cross-references: GB:M93676; NID:9212448; PIDN:AAA48985.1; PID:9212449

A>Note: sequence extracted from NCBI backbone (NCBIN:112864)

C:Comment: Alternatively spliced segments 1 and 2 are found exclusively in nonmuscle

C:Superfamily: myosin heavy chain; myosin motor domain homology

C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; myosin

F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <M9N>

F:1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <

F:88-802/Domain: myosin motor domain homology <M9OT>

F:178-185/Region: nucleotide-binding motif A (P-loop)

F:212-221/Region: alternatively spliced segment 1 #status experimental

F:559-593/Region: actin binding #status predicted

F:632-652/Region: alternatively spliced segment 2 #status experimental

F:692-714/Region: actin binding #status predicted

F:875-2007/Domain: coiled coil #status predicted <COI>

F:875-1315/Region: S2

F:1316-2007/Region: light meromyosin

F:129/Modified site: N6,N6-trimethyllysine (Lys) #status predicted

F:184/Binding site: ATP (Lys) #status predicted

F:132,742/Active site: Cys #status predicted

F:1954/Binding site: phosphate (Thr) (covalent) #status predicted

F:1987/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 6.4%; Score 205.5; DB 1; Length 2007;
Best Local Similarity 20.1%; Pred. No. 0.5;
Matches 159; Conservative 129; Mismatches 291; Indels 211; Gaps 33;

11 GPESVSSNOSSMNP I--INGQINSEFETKESPTA-----EASPS 48
DB 1130 GDEAVQKNNKALVITELQAIQLEQDELESEKASNNKAEKQKRDISELEALKTLEQRT 1189
QY 49 ASSSVSSWFLSSAKNALISLRDALINKSSPTDSLSQLEASTSTSTVTVAAKDYDEA- 107
DB 1190 LDTTAAQOELRTKREQEVAELKKAIEETKKNHQAQOETIRRHATA--LEELSEQLQAK 1247
QY 108 --KSNPDYAKSGLEN-----AKTL-----AYETKMDLMAALQDME-----RL 144
DB 1248 RFRANLEKKNQGLSESNKELACEVYKLAQVKAISEKRRKLLAQVQELTAKYTEGERLAV 1307
QY 145 ANSPDPSNNHTEEVNNTKKAL-BAOKDTIDKLVTLQONKSLTFTVL-KTTDSADQIPA 202
DB 1308 ELAEKANKLQNELDNVSSLEPEAKKGIKFAADAASLEQDQDTQLELDETRQKNTLS 1367
QY 203 INSQLEINKNS-----ADQIKDLERONISYEAVLTNAGE-----VIRASSEAGIK 248
DB 1368 RIQLEEEKNNLQEQEEREEAKRNLEKQMLALQALAAKAKVDDDLGTIEGLEENRKK 1427
QY 249 LQALQDST-----VQAGQDSQ-AAVLAQAOQNNSPN 278
DB 1428 LKQDMESLSQRIEEKAMAVDKLEKTKNRLQOELIDMLVLDHQRIQVSNLEKQKKFEDM 1487
QY 279 IATKELI-----DAETKYVELKQEHGTGLTDSPLVKA-----EEQISOAQKDIO 324
DB 1488 LAEEKNIISARYAEERDRAEARE--KETKALSLAALAELEAKFEFERQKKQLRADME 1545
QY 325 EIKPSGSDIPYVPGSGSAASAGALKSSNNS-----GRISLLIDVDYDNEMA----- 374
DB 1546 DLMSSSKDQV-----GKNVHELKSKRTLEQOQVEEMTQLELEDELOATEDAK 1593
QY 375 ----IALQFRSMIEQFNVPNPATKELQAMEQLTAMSDQVVGANGELPAETQAKD-- 428
DB 1594 LRLEVNQAMKQFE-----RDLOARQDQNEKKRMVLYQVRELEAELEDERKOR 1643
QY 429 ALAQAQKQSDADLATAMGQVAAKAVGGSGAGTACTYQNMVKKQYKTAFFSTSSSYA 488
DB 1644 ALAAVAAKKKMEHDKRLQEOI--EAAANKARDEA-----IKQARKL---QAQKKQDQ 1689
QY 489 AALSDGYSAVKTLSNLYSESR-----SGVQSAISQJANP-ALSRSVSRSGIESQGRSAD 541

```

Db 1690 RELEE---ARASRDEIFAQSEKSEKKLGLAEIILQLOEPFASERARRHAEQERDLAD 1746
Oy 542 ASQRAEETIVRSOTLGDVYSRLQVLDLSMTIVSNPQANOE---EIMOKLTASISKAPQ 598
Db 1747 EIANAS---GKSALLDKRRRLERARIAQLEBELEEOOSNMELNFRFKTLQV----- 1797
Oy 599 FGYPVONSADSLQKFAAOLEREFFVGDERSLAESOENAFRKO-----PAFIOQVLVN 650
Db 1798 -----DTLNSELA-----GERSAAOKSENA-RQOLERONKELAKKLOLEGS 1838
Oy 651 IASLESGYLS 660
Db 1839 VKSKFKATIS 1848

```

RESULT 30

```

T18351
Imp1 protein - Mycoplasma hominis
C:Species: Mycoplasma hominis
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 07-Dec-1999
C:Accession: T18351
R:Jensen, L.T.; Ladefoged, S.; Birkelund, S.; Christensen, G.
Infected. Immun. 63, 3336-3347, 1995
A:Title: Selection of Mycoplasma hominis Pg21 deletion mutants by cultivation in the pre
A:Reference number: Z18884; MUID:95369882; PMID:7543881
A:Accession: T18351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1051 <JEN>
A:Cross-references: EMBL:U21961; NID:g790241; PID:g790242; PIDN:AAA81012.1
C:Genetics:
A:Gene: Imp1
A:Genetic code: SGC3

```

```

Query Match 6.4%; Score 204; DB 2; Length 1051;
Best Local Similarity 19.4%; Pred. No. 0.24;
Matches 139; Conservative 142; Mismatches 315; Indels 122; Gaps 30;

```

```

Oy 5 HHMHSGPSSVSSNOSSNMPI-----INGQIASNSETR---ESTKASASPSASSSVSS 55
Db 144 YEHK---QNIENELNKYTPISLESTLIEIQNATNMLIKLNESTREKKNIDNLAKEQL 200
Oy 56 WSFLSSAKNAL--ISLRDAILNKNSSPTD-----SLSQLASTSTSTVTRAAKDVDEAKS 109
Db 201 KASISQANQLPOLSDNDSSEIAKAKSLDAETKMANQAVASNTASM-----QSAS 252
Oy 110 NEDTAKSGLENAKTLAEY---ETKMAADLAALODMERLANSDPN-NHTEEVNNIKKAL 164
Db 253 SLDAKYA--ETKKTLETFNKDKAEKFNELKQTRNOIOEFINTNKNPNYSELISQUTSKR 310
Oy 165 ENOKRTIOKLNKLVTLQONKSLTEVLTQTTSDADQIPAINSOLEINKNSADQITDLEQ 224
Db 311 DSKNSVTSSNK-SDIESNTEFLKQALAKA-NADRVQADN---LAKSIRKQLNNSVSA 364
Oy 225 NISYEAULTNAGEVIKASSEAGIKLQALQISVDAGDOQAVALQAOQNSPDNIATKE 284
Db 365 N-TLSAKLTDRKNTTO---QAKTELEKEIQKANQAIKSNMTASMOGAKSLAKVAELTK 420
Oy 285 LIDA---AETRVNELKQEHNTGLTDSPLVKAEEQISQAQKDIQETIKPSGSDIPVPGSG 340
Db 421 KLETENKDKAEKFNELKQTRNOIOEFINTNKNPNYSELISQUTSKRDKNS---VTDS 477
Oy 341 SNASGSAAGALKSSNNSRISLLDDVDVNEMAALALQCFRMTIOF---NVNPNATK 396
Db 478 NISDIESNTEFLKQALAKAKA-KSSIDNELRPK-NDLOSRIEFGIRRTNTFSWISS 534
Oy 397 EIQAMEAOLITAMSDQVGD---GELPAETQAIKDALQALK--QPSADGLATMAGOVAF 451
Db 535 KLETTKKNL---AELITRKDAIKNPNSSKQALKDSSQOVQKLGNELLTITIEEFGKVT 591
Oy 452 AAAYGVGSAGTAGTVQNM-----VKOLY-----KTAFFSTSSSSVAALSDGYS 496

```

```

Db 592 KNSNIGYRLFLKLAQEQFNNSDVDKLKNAMEEQKTLSSKKQKIGNOSTKDYLTQLSTEMS 651
Oy 497 AY-----KTLNLSYSESRGCVOSAIQSOTANPLNSV-----SRSGIESGGRADASQRA 547
Db 652 TQESTIKRYIVMIOAHIRNNLSQYRLKADKLIANNKRGYGDVGIESLQKQDMLD--- 708
Oy 548 ETIVRDSOTLGDVYSRLQVLDLSMTIVSNPQAN-----OE 583
Db 709 DSVLSYDSDSLKDFNK--ALRVLVGDYTKNPNVSSWFIKRRKRSTIENTONLNLIVRPN 766
Oy 584 EIMOKLTASISKAPQFGYPVONSADSLQKFAAOLEREFFVGDERSLA---ESQENAF 637
Db 767 EILDKAKDLDRAEKTIKFVDENINSLDQRAKIRKQETILANKDLSNFTLNHOKNQF 824

```

RESULT 31

```

MKMKW
myosin heavy chain B [similarity] - Caenorhabditis elegans
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 13-Jun-1983 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C:Accession: T20770; T21629; A93958; A93287; A21074; A02992
R:Kershaw, J.
Submitted to the EMBL Data Library, November 1996
A:Reference number: Z19322
A:Accession: T20770
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1963 <WIL>
A:Cross-references: EMBL:Z81499; PIDN:CAB04089.1; GSPDB:GN00019; CESP:F11C3.3
A:Experimental source: clone F11C3
A:Accession: T21629
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1963 <W12>
A:Cross-references: EMBL:Z83107; PIDN:CAB0505.1; GSPDB:GN00019; CESP:F11C3.3
A:Experimental source: clone F13A7
R:Karn, J.; Brenner, S.; Barnett, L.
Proc. Natl. Acad. Sci. U.S.A. 80, 4253-4257, 1983
A:Title: Protein structural domains in the Caenorhabditis elegans unc-54 myosin heavy
A:Reference number: A93958; MUID:83273600; PMID:6576334
A:Accession: A93958
A:Molecule type: DNA
A:Residues: 1-61, 'EMSVIO', 65-376, 'V', 378-1963 <KAR>
A:Cross-references: GB:V01050; NID:g156399; PIDN:AAA28124.1; PID:g156400
R:Macchiann, A.D.; Karn, J.
Nature 299, 226-231, 1982
A:Title: Periodic charge distributions in the myosin rod amino acid sequence match cr
A:Reference number: A93287; MUID:82272395; PMID:7202124
A:Accession: A93287
A:Molecule type: DNA
A:Residues: 847-1333, 'R', 1335-1876, 'L', 1878-1963 <KCL>
R:Wills, N.; Gesteland, R.F.; Karn, J.; Barnett, L.; Bolten, S.; Waterston, R.H.
Cell 33, 575-583, 1983
A:Title: The genes sup-7 X and sup-5 III of Caenorhabditis elegans suppress amber non
A:Reference number: A21074; MUID:83232892; PMID:6571695
A:Accession: A21074
A:Molecule type: DNA
A:Residues: 1873-1963 <W13>
A:Cross-references: GB:V01494; GB:J01049; NID:g6783; PIDN:CMA24738.1; PID:g6784
C:Genetics:
A:Gene: unc-54; CESP:F11C3.3
A:Map position: 1
A:Intons: 21/3; 64/3; 111/3; 264/1; 525/3; 951/2; 1747/3; 1819/3; 1894/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle
F:84-775/Domain: myosin motor domain homology <MKOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:662-684/Region: actin binding #status predicted
F:766-780/Region: actin binding #status predicted
F:848-1963/Domain: coiled coil #status predicted
F:848-1162/Region: S2
F:1163-1963/Region: light meromyosin

```


A:Molecule type: mRNA
A:Residues: 2035-2116
R:Magle, G.; Noegel, A.; Scheel, J.; Gerisch, G.
FEBS Lett. 227, 71-75, 1988
A:Title: Phosphorylation of threonine residues on cloned fragments of the Dictyostellium
A:Reference number: S00250; MUID:88112226; PMID:2828113
A:Accession: S00250
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1734-1893 <NAG>
C:Comment: The rod domain is highly periodic, containing a pattern of 7-residue repeats
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolyase; nucleotide binding; P-loop; phos
F:1-818/Domain: globular head <HED>
F:89-747/Domain: myosin motor domain homology <MMOT>
F:179-186/Region: nucleotide-binding motif A (P-loop)
F:819-2116/Domain: alpha-helical rod <ROD>

Query Match 6.3%; Score 201; DB 2; Length 2116;
Best Local Similarity 20.7%; Pred. No. 0.83;
Matches 163; Conservative 130; Mismatches 313; Indels 182; Gaps 33;

Oy 9 ESGPESVSSNOSSMNPINGOIASN--SETKESTKAS--EASPSASSSVSSMSF--LSS 61
Db 756 EAREORISITITAIQOATRGWJARVYVKOAREHTVAARITQOQLRAYIDKSWPMWKLFS 815

Oy 62 AKNALISLND-----AILNKNSSPTDSLQ--LEASTSTVTRV-----A 100
Db 816 KARPPLKRRNFEKEIKERELELLEKSNLTDSTQDKLEKSLK-DTESNVLDLQRLKA 874

Oy 101 AKD-----YD-----EAKSNFDTAKSGLENAKT-LAEYETKMDLMAA 137
Db 875 EKETLKAMWSDKALEAQKRELEIRVEDMESELEDEKLLLENLQOKRVSVEEKVRDLREE 934

Oy 138 LODMERLANS--DPSNNHTEEVNNTKKALEAQKOTIDKLNKLV-ILQONNKSLTE----- 189
Db 935 LOEBOKLRTLEKLLKKKEELEEMKRVNDGSDPTISRLEKIKDELQOKVEELETSFSEE 994

Oy 190 -----VLKTTDSADQIPAINSOLEINKNSAQOIIKOLEPONTISYEAVLTNAGEVIKASSE 244
Db 995 SKDKQVLETRVRLQSELDLTVRLDSETKDK--SELLRQKKKLEELKQVDEALAEFLA 1052

Oy 245 AGI-----KLGQALQSLVDAGD-----QSOAAVLAQOONNSPDNIAA 281
Db 1053 AKLAQEAANKRLQGEYTELENEKFNSEVTPARSNVEKSKTLESQVAVNNLEDEKKNRDA 1112

Oy 282 TKELIDAEETKYNELKQ--EHTG-----LTDSPLYKKAE-----EQISAQOQDIOETIKS 329
Db 1113 LEKKRKALDAMEKMDQESTGCKSLYDLKVKQESDMELARNQOISELOSTIAKLEKI 1172

Oy 330 GSDIPIVPGSGSAA--SAGSAGALKSSN--NSGRISLLDD-----VDNEMAAIALQG 379
Db 1173 KSTL-----EGEVARLQGLELEAQAKSNVEKQKKVELELDLQKSAQLAEEFLAAQALDQ 1227

Oy 380 FRSMTIQ-----FNVNNPATAK-----ELOAMEAQLTAMSQDLVG 414
Db 1228 LKKKLEQELSEVQTOLESEANNKNVNSDSTNKHLETSFNNLKLELEAQAKALEKKRLG 1287

Oy 415 ADGELPAEIQAKDALQAKOPSD-----GLATAMQV--AFPAAVVGGGSACT 463
Db 1288 ---LESELKHNVEOLEEKKQKESNEKKRVLEKEVESLQDQIEEVASAKVAETEAQKK 1343

Oy 464 AGTVQNNVKOLKTFAPSTSSSSSYAALSDGYSAKVTNLSLSESGVQSAISOTANPA 523
Db 1344 KESELDKTRQYADVSSHKDS-----VEQLKTLQAKKEELRNTAEELGQDLRAE 1394

Oy 524 LRSVSRSGIESQGRS-----ADASQRAETIVRDSQT-LGD---VYSRLQVLDLS 569
Db 1395 RSKKKAEPDLBEAVKNLEETAKKVAEKAMKKAETDYKSTKSELDQAKNVSSQYVOLK 1454

Oy 570 LMSTIVSNQANQDELMQKLTASI--SKAPQFCYPAVQNSASDLQFAQOLEREVEDGER 627
Db 1455 RLNEELSELRLVLEADERCNSAIKAKTAEASALESLADEIDAANNAKAKAEKRSKELEV 1514

Oy 628 SLAESQEN 635
Db 1515 RVAELEFS 1522

RESULT 34
B90835
probable tail fiber protein [imported] - Escherichia coli (strain 0157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90835
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gatawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and 9
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90835
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835073.1; PID:913361114; GSPDB:GN00154
A:Experimental source: strain 0157:H7, substrain RIMD 0509952
A:Gene: Ecs1650

Query Match 6.2%; Score 200; DB 2; Length 971;
Best Local Similarity 20.7%; Pred. No. 0.32;
Matches 121; Conservative 84; Mismatches 219; Indels 160; Gaps 17;

Oy 6 HIMESGPESVSSNOSSMNPINGOIASNSETKASEPASPASSSVSSWFLSSARNA 65
Db 144 HATDAADSAARAASAGQASASAGSAGTATKATPASKASAAAESSKGAATSAAGA 203

Oy 66 LISLDAILNKNSSPTDSLQLEASTSTVTRVAAKDYDEKSNFDTAKSLENAKTLA 125
Db 204 A-----KTSETNAVVSQOASATSTATTAKS---EASSARDAASAKAEAKS-- 248

Oy 126 EYETKMDLMAALQOMERLANDPSNNHTEEVNNTKKALEAQKOTIDKLNKLVTLQONNK 185
Db 249 -SETSNAS-----SASSAASSATAGNSAKAA----- 274

Oy 186 SLTEVLKTTDSADQIPAINSOLEINKNSAQOIKDLERONISYEAVLTNAGEVIKASSEA 245
Db 275 -----KTSE-----TNKSSSETAAEQ-----SASAAA 296

Oy 246 GIKLQALQSLVDAGDQSOAAVLAQOONNSPDNIAATELIDAEETKYNELKQERTGLTD 305
Db 297 GSKTAAALASASASTSAGQASASATAGKSASASAS--ASTATTGAGEATEQASAAAS 353

Oy 306 SPLVKAEEQISOAQKDIOEIKPQSGSDPIVPGSGSASAGSAGALKSSNNSGRISILL 365
Db 354 SASAAKTSETNAKA-----SETSABESSKTRAAASSASASSASASASAK----- 397

Oy 366 DDVDNEMAAIALQGRSMIEDPFVNNNPATAKELQAMEAQLTAMSQDLVGADGELPAEIQ 425
Db 398 DEATRQASA-----AKSSATVTASTKATB-----A 421

Oy 426 IKDALQALQKOPSDAGLATAMQOVAFAAAKVGGSQA-GTAGTVQNNVKOLKTFAPSTSS 484
Db 422 AGSATPAALQOSKSTASATRAETAKKRAEDIASAVALDEASTTKKGIYOL--SSATNSTS 479

Oy 485 SSTAAALSDGYSAYVTNLSLY--SESGVQSAISOTANPALSRSVSRSGIESQGSADAS 543
Db 480 ESTLAATPPKAVKAYELANGKTTAQAATTAQOKGIYOLSNATNSTSEMLA-----ATRS 532

Oy 544 ORAETIVRDSQTLGD-----VYSRLQVLDLSLMSTIVSNQ 580
Db 533 VAAAVDLANGKYYTAQDATTAQOKGIYOLSSATNSASETTLAATPKA 576

RESULT 35
C85693

probable membrane protein of prophage CP-933x Z1918 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: C85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
M.L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <STO>
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWGP:Z19
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1918

Query Match 6.2%; Score 200; DB 2; Length 973;
Best Local Similarity 20.7%; Pred. No. 0.32;
Matches 121; Conservative 84; Mismatches 219; Indels 160; Gaps 17;

QY 6 HHMSGESEVSSNOSMNPLINGQIANSSEKSTKASESPASSSSVSMSFLSSAKNA 65
DB 146 HATDADSRARASTSAGQAASSAQSSASGASTKATERSKAAAESSKSAATSAQA 205
QY 66 LISLRDALINKSSPPTSLSOLEASTSTSTVTVRAAKDYDEAKSNFDTAKSGLENAKTLA 125
DB 206 A-----KTSETMAAVSQSSAATASATATTKAS---EASSARADASSEAKS 250
QY 126 EYETKADLMALQDMERLANSNHTSEVNNIKKALEAKQDITDKLKVLTIONOK 185
DB 251 -SETSAAS-----SASSAASATPAAGNSAKAA----- 276
QY 186 SLEVEVKTETSDAQIPAINSOLEINKNSADQITKDERONISYEAVLTNAGEYIKASSEA 245
DB 277 -----KTSE-----TNKSSSETAAEQ-----SASAA 298
QY 246 GIKLGALQSIYDADQSOAAVLAQOONNSPDNIAATKELIDAETKVELKOEHTGLD 305
DB 299 GSKTAAALASASASTAGSASATPAAGKSAEBSAAS---ASTATTKAGEATQASAAAS 355
QY 306 SPLVKAEEQISOAQKDQIEIRPGSGDPIVGPSSGAASAGSAGLAKSNNSGRISLL 365
DB 356 SASAATSETNKA-----SETSAESSKTAASASSASSASSASASK----- 399
QY 366 DVDNEMAAIALQGFMSIEGFVNNPATAKELQAMEAQLTAMSDQLVGADGELPAEIOA 425
DB 400 DEATRQASA-----AKSATTASTKATE-----A 423
QY 426 IKDALAQLKQPSADGLATAMGOVAPAAKVGGSAG-AGTIVQNMVKQLYKTAFFSSTS 484
DB 424 AGSATPAASQSKSTAESAAETPAARADIASAVALEDSATTKGIYQL--SSATNSIS 481
QY 485 SSSAAALSGCYAKYKTLNLY--SESSGVSQSAISQANALNSVSRSGIESGGRADAS 543
DB 482 ESILAAPKAVKAYELANQKTYAQDATTAKGIYQLSNATNSTSEMLA-----ATPRS 534
QY 544 QRAAETIVDSQITGD-----VYSRLQVLDISMTISTVSNPOA 580
DB 535 VRAAYDLANGKTYAQDATTAKGIYQLSSATNSASSETLATPRA 578

RESULT 36
S02771
myosin heavy chain A [similarity] - Caenorhabditis elegans
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Caenorhabditis elegans
C:Date: 31-Dec-1993 #sequence_revision 19-May-2000 #text_change 19-Apr-2002
C:Accession: T23622; S02771
R:Harris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19773

A:Accession: T23622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1992 <MIL>
A:Cross-references: EMBL:Z78199; PIDN:CA801576.1; GSPDB:GN00023; CESP:K12F2.1
R:DiDio, N.J.; Maruyama, I.N.; Krause, M.; Karn, J.
J. Mol. Biol. 205, 603-613, 1989
A:Title: Sequence analysis of the complete Caenorhabditis elegans myosin heavy chain
A:Reference number: S02771; MUID:89178677; PMID:2926820
A:Accession: S02771
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-1116,140-1992 <DIB>
A:Cross-references: EMBL:X08067; NID:g6798; PIDN:CAA30856.1; PID:g6799
C:Genetics:
A:Gene: myo-3; CESP:K12F2.1
A:Map position: 5
A:Introns: 46/1; 192/1; 292/1; 468/2; 1921/3
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; muscle
F:89-802/Domain: myosin motor domain homology <MMOT>
F:202-209/Region: nucleotide-binding motif A (P-loop)
F:690-712/Region: actin binding #status predicted
F:793-807/Region: actin binding #status predicted
F:875-1189/Region: coiled coil #status predicted <COI>
F:1190-1992/Region: light meromyosin
F:153/Modified site: N6,N6-trimethyllysine (Lys) #status predicted
F:208/Binding site: ATP (Lys) #status predicted
F:730,740/Active site: Cys #status predicted

Query Match 6.2%; Score 200; DB 1; Length 1992;
Best Local Similarity 18.9%; Pred. No. 0.84;
Matches 142; Conservative 152; Mismatches 304; Indels 154; Gaps 28;

QY 28 GQIASSEKSTKASESPASSSSVSMSFLSSAKNAL---ISLRDALINKSSPPTS 83
DB 886 GELAVKIQLE--EAVQRELIARSOLESOVADLVEKNALFSLTERKANLADAERNEK 943
QY 84 LSOLEASTSTVTVRAAAYDYDEAKSNFDTAKSGLENAKTLAEYETKMDLMAALODMER 143
DB 944 LMDLRA-TLESKLSDTGQLEDMQERNEDLAROKKKTQDELDTKKHYVDLSELKRAEQ 1002
QY 144 -----LANSDFS-----NNHTEVY-NIKKALEAKQDITDKLKL 177
DB 1003 EKOSRDHNIIRSLQDEKANDQEAVALNKREKKHQQEENRKLINEDLSQSEDKVNHLEKIRNK 1062
QY 178 -----VTIQONOKSLTEYK-----TTSDADQIPA----- 202
DB 1063 LEQMDLEENIDREKRSGLIEKAKRVEGDLKVAQENIDETTRQKHDEVTTLRKEED 1122
QY 203 ---INSOLEINK---SABQITKDERONTSEAVL---TNAGEYIKASSAGIKLQGA 252
DB 1123 LHHTNAKLLENNSITAKLORLIKELTARNAELEEELFAERNRSKSDRSRLAELEEL 1182
QY 253 LQSIDAGQSOAAVLAQOONNSPDNIAATKELIDAA--ETKVELAQER---TGLTD 305
DB 1183 TERLEQOGG-ATAAQLEANKKREAEIARLREKEEDSLHETAFISLRHGDVAEELTE 1241
QY 306 -----SPLVKAEEQISOAQKDQIE-----IKPSGSDITVGPS 339
DB 1242 QLETLOLKAKESEAKSKLQRDLESQHAIDSEVSRQDLEKALKTIEQVSELTQKADE 1301
QY 340 GSAAASGAAGALKSSNNSGRISLLDQVDNEMAI--ALQGFMSIEGFVNN----- 391
DB 1302 QSRQLODFPALNRLNNEISDLNRSLEEDNDLNSLHRLKSTLQSLDETNNYDEBSRE 1361
QY 392 ----PATAKELQAMEAQLTAMSDQVAGDELPAEIOAIKDALQALQPSADGLATANG 447
DB 1362 RQALATATKANLEHNTILREHLEDEAKSKADLTROIQSKMAETIQQMKARFQSEGL-NMLE 1420
QY 448 QVAPAAKVGGSAGTAGTIVQNMVKQLYKTAFFSSTSSSSVYAAALSDGYSAYKTLNSIYSE 507

```

Db 1421 EIE-AKK-----ALQKLVDEL-----TDTEGCLFAKIASOEKVRFKIMQDL-DD 1463
OY 508 SSGVOSATISOTANPALSSVSSSGIESGGRSADASQRAETIVRSQTLG-DVYSRLQY 566
Db 1464 AOSDVEKAAQAQVAFYEHKRRQEPESITAEKKKTKTDLSSELDAQAQRNRLSTDLFRAKTA 1523
OY 567 LDSLSTIYSNPOANE--EIMOKLTASISKAPQFGYPAVQNSADSLQKFAAQLEREFD 624
Db 1524 NDLAEYLDSTREKNSLAQEVKDLTDQEGEGR-----SVALLQKIVKRLVEKEE 1575
OY 625 GERSLAESQENAFKQAPFIQOVLVNIASLFS 656
Db 1576 LQKALDEA-EAALAEAEAKVLRAQIEVSQIRS 1606

RESULT 37
C97038
phage-related protein, yqpo B, subtilis homolog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: C97038
R:Noilling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clc
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: C97038
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2052 <KUR>
A:Cross-References: GB:AE001437; PIDN:AAK79094.1; PID:q15024039; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1120

Query Match 6.2%; Score 200; DB 2; Length 2052;
Best Local Similarity 18.8%; Pred. No. 0.88;
Matches 137; Conservative 136; Mismatches 289; Indels 166; Gaps 26;

OY 26 INQIANSSEKSTKASPSASSVSSWFLSSAKNALISLDAIINKNSPDSIS 85
Db 743 LRKAVAGTMTBE--EASALEAMQIKM-----KATVSAEBA-----SSANNALS 786
OY 86 QLEASTSTSTV-----TVAAKD-----YDEKSNPDIAKSGLEAKKTLA 125
Db 787 VAKRSATLSTGCLNALGSIYVAGLTIGLSMAISDIIEKHEDKQKVDKLQGYE----- 840
OY 126 EYETKADLMAALQDMERLANSPPSNHTEEVNNIKKALEAQKDTIDKINKLVTLOQNK 185
Db 841 -----DLTKAMKONSITADNDYKNLEKEOSTLENALKRRALEGGKINSVKKSKKDK 893
OY 186 ---SLTEVLKTTDSADQIPAINSOLEINKNSAD-----QITKLERONISYE 229
Db 894 WGNSTLDDTKLDFOSELKNVKNVISETQKNLDTGTGYSDKTTGKIKLSEASEQJANNK 953
OY 230 AVLTVNAGEVTKASSEAGITGQALOSTIVADGDSQAALVLAQONNS-----PDNIA 280
Db 954 STBGAARDLTENSDDKAKTISIGLLIYSTVLYKOSDKTALSEQOKLAQALALVFQDLTL 1013
OY 281 ATKRELIDAAETRVNEL-KOEHTGLTDSPLVKKA---EOISQAKNDIOETKPSGSDIPV 336
Db 1014 STDKNGDTIINKNASALQKTEAFIKGNNAVVAAILNDKELQQAQEKINK----- 1062
OY 337 GPGSGAASGSAAGALKSSNSGRISLLDD-----VDNEKAATIALOGFNSMITEQFN 389
Db 1063 -----AIQSTTNSKYSLTYSNQIERNNKLDASVADIKRNHDELVPYLLNN 1108
OY 390 NNPAATAKELQAMEA--QLTAMS-DOLVGAD---GELPAEIOAKIALQALQALQ-OPSANG 441
Db 1109 TIKQLISKIOQGISFKNTYVVMSTAOQMADQKNYSKLGAEYQOYTYLKQKISKNKSSKE 1168
OY 442 LATAMQOVAFAAA---KVGGSAGT-----AGTVOMNV-----KOLYKT 477

```

```

Db 1169 LADAKQOVKIKIADLANIKGLSTVTDKNGNTYTLKAGNKEKEIKLIKSEGNTRYKELEKI 1228
OY 478 ARFSSTSSSSYAALSDGYA-----KTLNLSYSESRGQ---SATSOTANA 523
Db 1229 KUGSSKSTY--MIEGNTAVTQKOQVIRIGYTYEAKMIAAKAKVKNIASTLETIKAGQA 1285
OY 524 LRSVSRSGIESGGRSADASQRAETIVRDSQTLGPDVYSRLQ-VLDSLSTIYSNPOAN- 581
Db 1286 LTSKQAKNAKE-----RIKALQEIIDLKFAKSEIDKLFQADQSADSSNNTYDSQDIGY 1341
OY 582 -OEIMOKLTASISKAPQFGYPAVQNSADSLQKFAAQLEREFDGERSLAESQENAFKQ 640
Db 1342 ASEDAEAKKEAKKEAKKEAKKEQQAEBKAERORETORAKAEQAKAEQQRKSLERE 1401
OY 641 PAFIQOVL 648
Db 1402 KKHMEAL 1409

RESULT 38
S45781
Probable calcium-binding protein YBL047c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YBL0520
C:Species: Saccharomyces cerevisiae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Apr-2002
C:Accession: S45781; S50284; S45782; S39841; S37339; S42498
R:Goffeau, A.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; de Wergifosse, P.; van Dyck,
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45745
A:Accession: S45781
A:Molecule type: DNA
A:Residues: 1-961 <GOR>
A:Cross-References: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c
A:Experimental source: strain S288C
R:de Wergifosse, P.; Jacques, B.; Jonniaux, J.L.; Purnelle, B.; Skala, J.; Goffeau, A
Yeast 10, 1489-1496, 1994
A:Title: The sequence of a 22.4 kb DNA fragment from the left arm of yeast chromosome
NA-binding protein.
A:Reference number: S50284; MUID:95176707; PMID:7871888
A:Accession: S50284
A:Molecule type: DNA
A:Residues: 1-961 <DEF>
A:Cross-References: EMBL:X78214
A:Experimental source: strain S288C
R:Dubois, E.; El Bakoury, M.; Glandsdorff, N.; Messenguy, F.; Pierard, A.; Scherens,
submitted to the Protein Sequence Database, August 1994
A:Reference number: S45782
A:Accession: S45782
A:Molecule type: DNA
A:Residues: 579-1381 <DUB>
A:Cross-References: EMBL:Z35808; GSPDB:GN00002; MIPS:YBL047c
A:Experimental source: strain S288C
R:Scherens, B.; el Bakoury, M.; Vlerendeels, F.; Dubois, E.; Messenguy, F.
Yeast 9, 1355-1371, 1993
A:Title: Sequencing and functional analysis of a 32 560 bp segment on the left arm of
A:Reference number: S39824; MUID:94205266; PMID:8154187
A:Accession: S39841
A:Molecule type: DNA
A:Residues: 579-1381 <SCH>
A:Cross-References: EMBL:Z23261; NID:g313733; PIDN:CAA80797.1; PID:g313748
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SGD:EDL; MIPS:YBL047c
A:Cross-References: SGD:S0000143
A:Map position: 2L
C:Superfamily: yeast probable calcium-binding protein YBL047c; calmodulin repeat homo
C:Keywords: calcium binding; EF hand; transmembrane protein
F:167-199/Domain: calmodulin repeat homology <EPI>
F:560-576/Domain: transmembrane #status predicted <TMM>

Query Match 6.2%; Score 199.5; DB 1; Length 1381;
Best Local Similarity 20.4%; Pred. No. 0.54;

```


THIS PAGE BLANK (USPTO)